

WILD LAB Analytics : Outil d'aide au diagnostic médical





CONTEXTE

Nombre croissant
de patients

Pénurie de
personnel médical

Diagnostic tardif

Mise en place d'un
traitement adapté

Evolution de la
pathologie

Qualité de vie du
patient

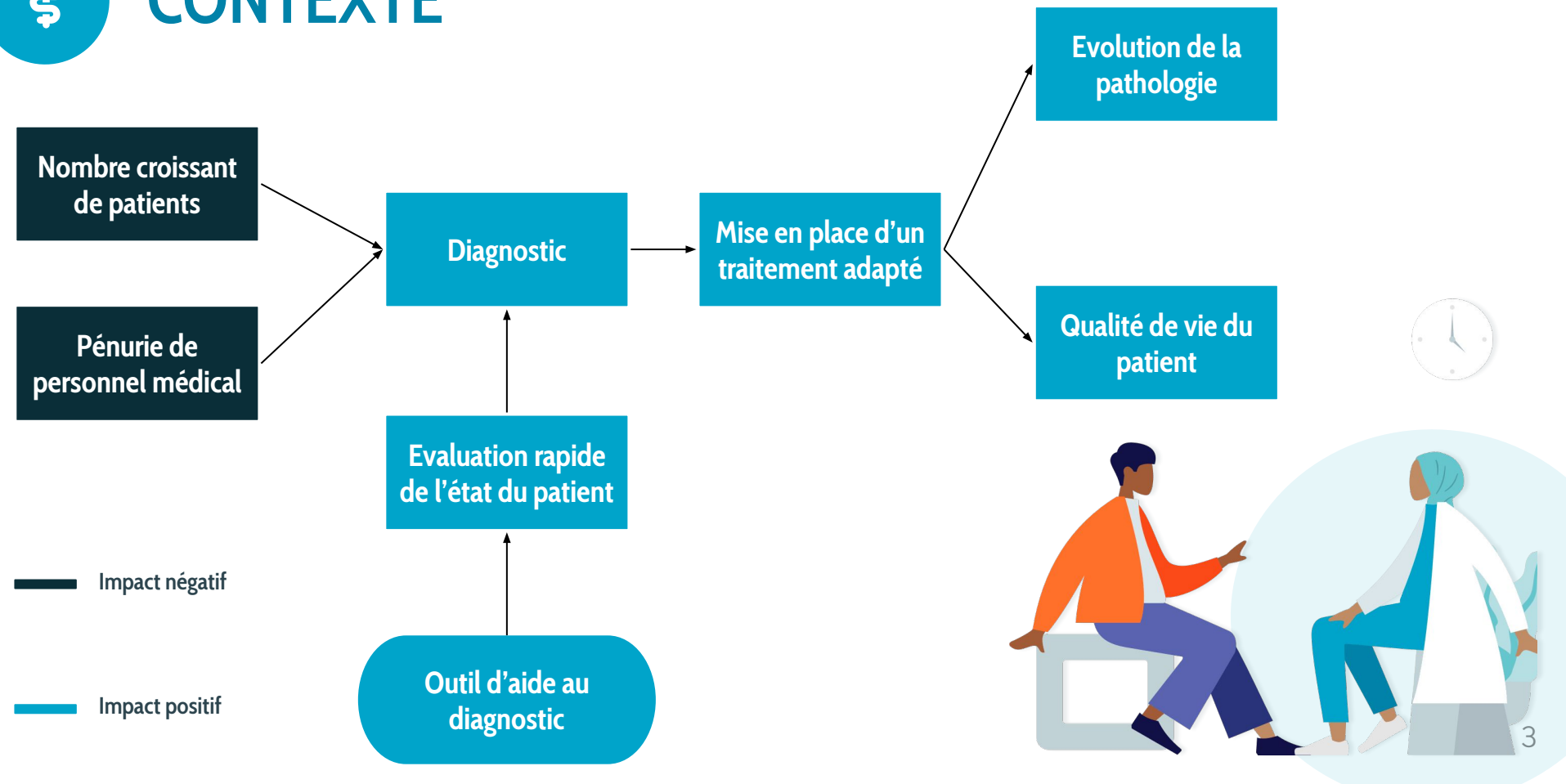
Impact négatif

Impact positif





CONTEXTE

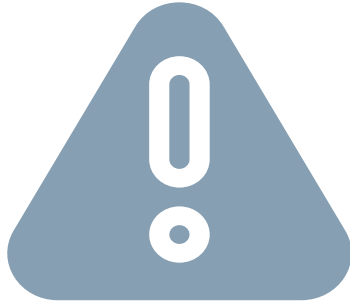




OBJECTIF



Prédire correctement les patients malades



Réduire au maximum la possibilité de prédire qu'un patient soit sain alors qu'en réalité il est malade





DOMAINES D'EXPERTISE DE L'OUTIL



Diabète



Cancer du sein



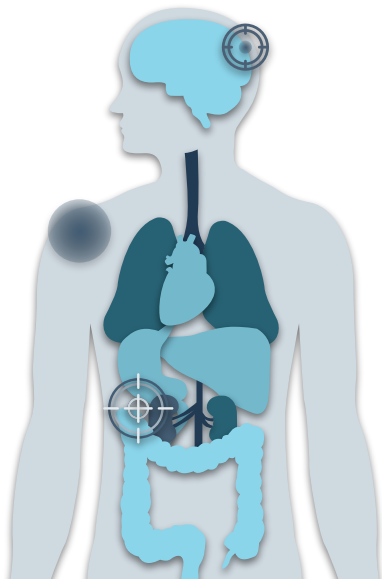
Maladies du foie



**Maladies rénale
chroniques**



**Maladies cardiaques
chroniques**





PROCESSUS DE CRÉATION DE L'OUTIL



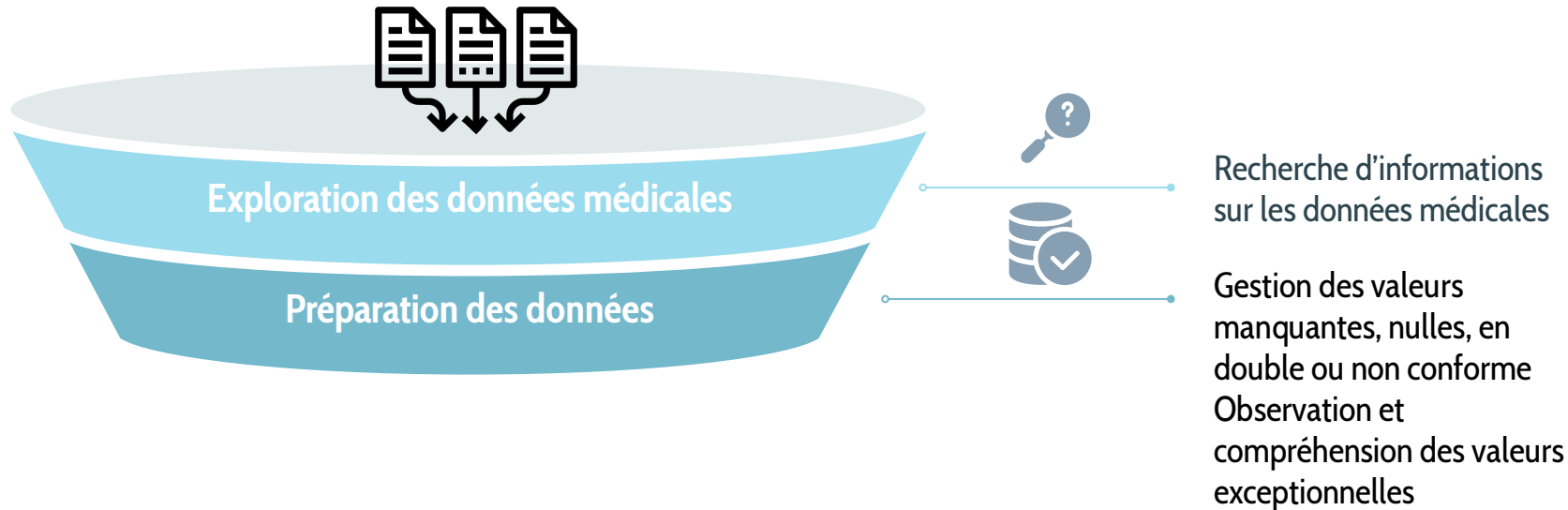
Exploration des données médicales



Recherche d'informations
sur les données médicales

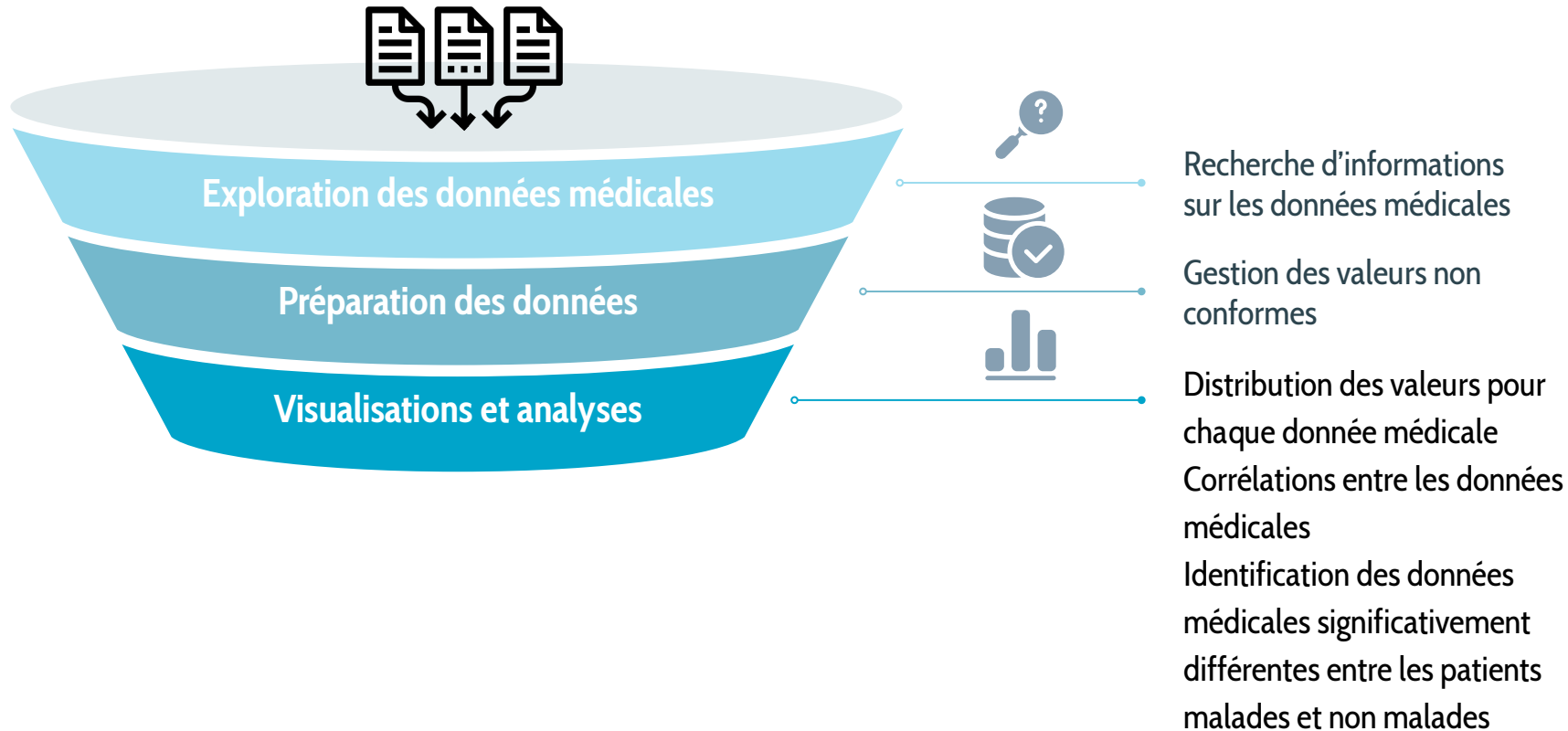


PROCESSUS DE CRÉATION DE L'OUTIL



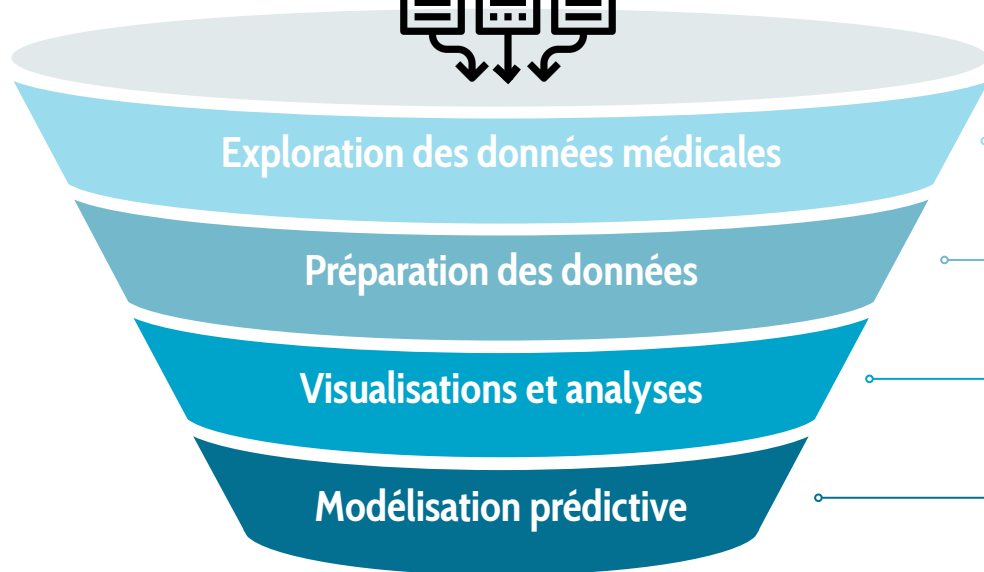


PROCESSUS DE CRÉATION DE L'OUTIL





PROCESSUS DE CRÉATION DE L'OUTIL



Recherche d'informations
sur les données médicales



Gestion des valeurs non
conformes



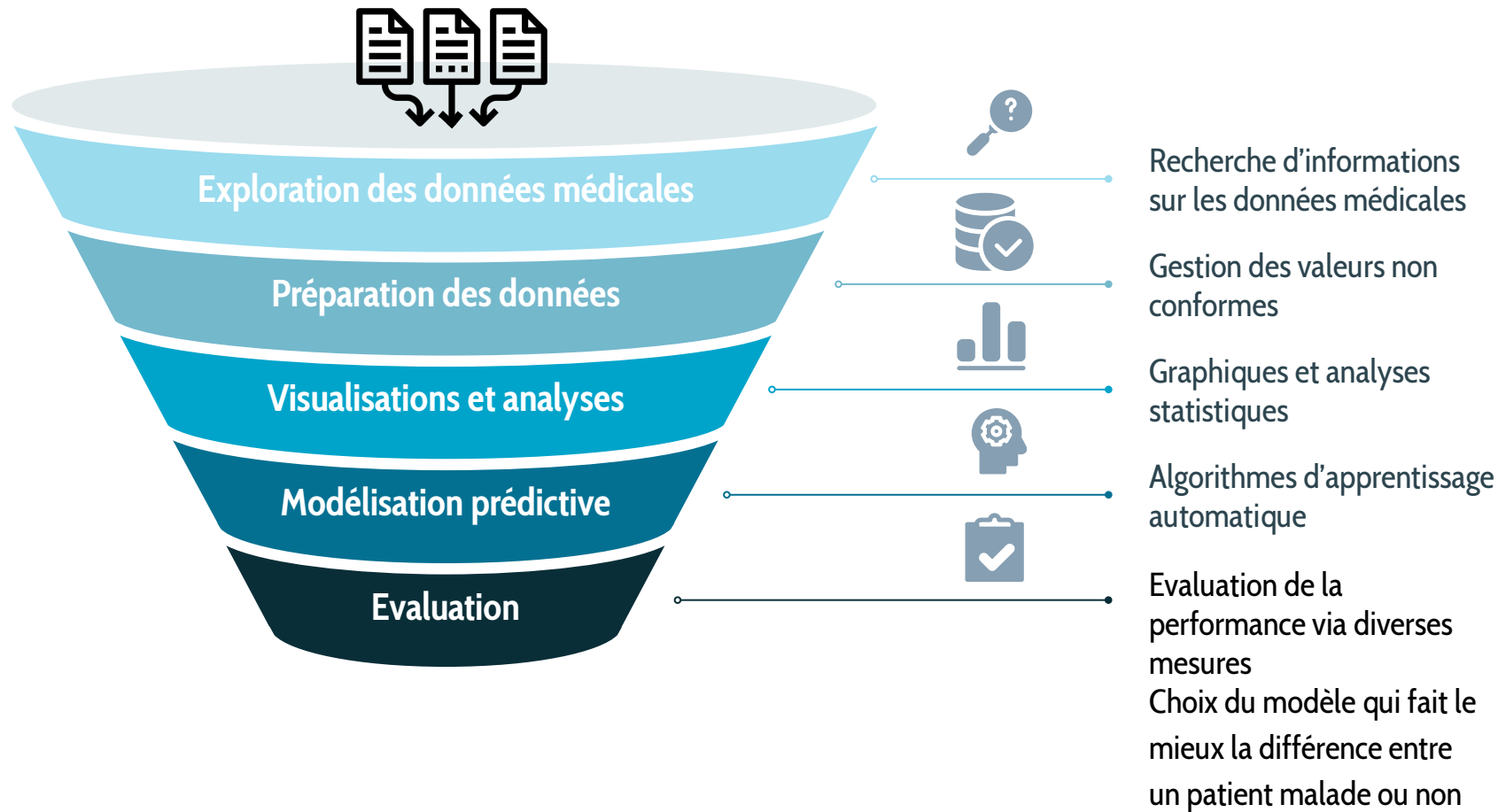
Graphiques et analyses
statistiques



Construction d'algorithmes
d'apprentissage
automatique pour prédire
si un patient est malade ou
non à partir d'une liste de
plusieurs modèles

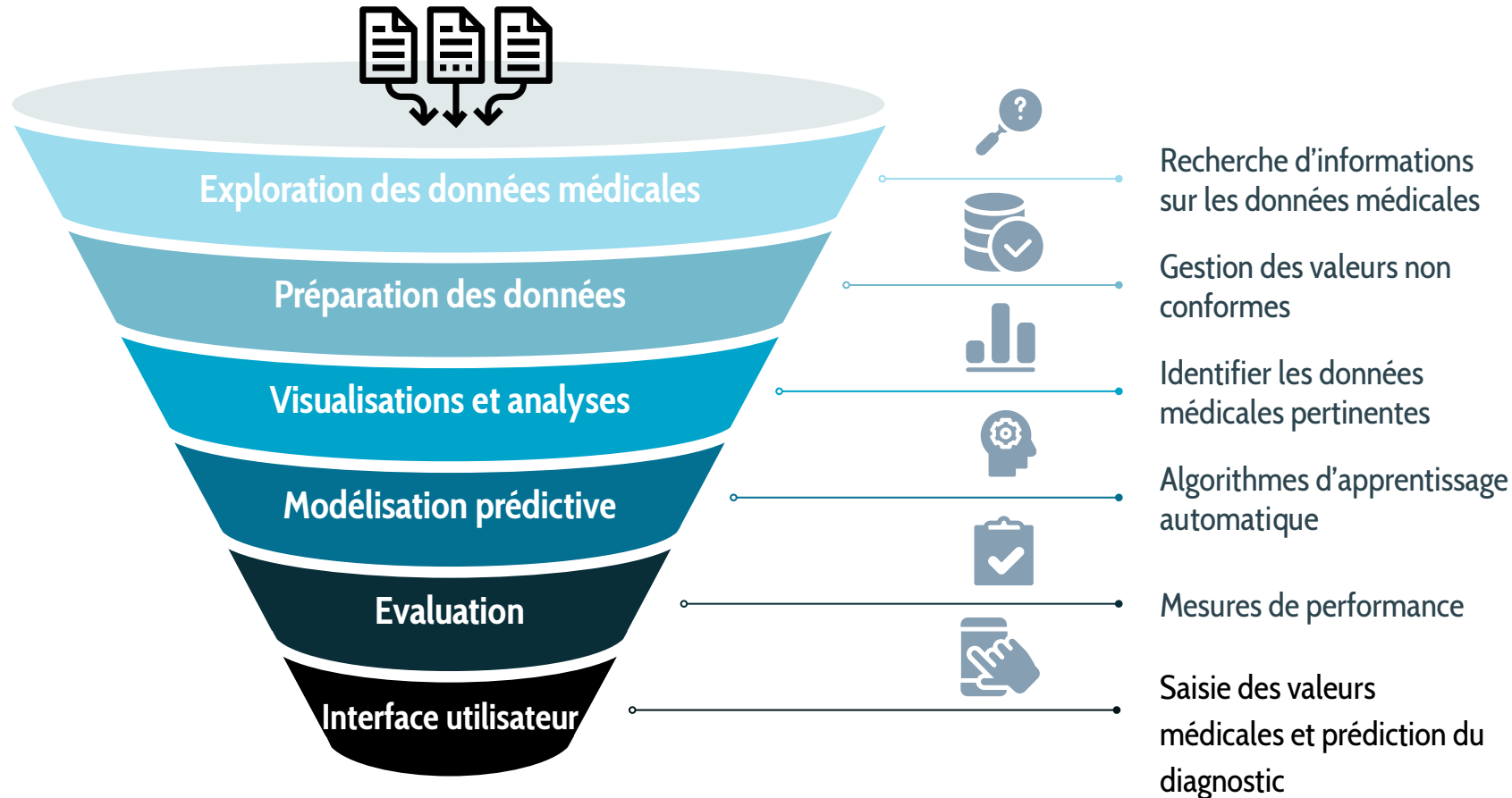


PROCESSUS DE CRÉATION DE L'OUTIL





PROCESSUS DE CRÉATION DE L'OUTIL





Technologies utilisées



Exploration des données : Python et ses librairies (Pandas, Seaborn, PlotlyExpress, NumPy)



Préparations des données : rééquilibrage des datasets avec RandomOverSampler



Analyses statistiques : SciPy et Statannot



Machine Learning : Cross-validation, RandomSearch, GridSearch, Scikit-learn, Keras, LASSO, SHAP







Application Web : Dash, Docker, Render

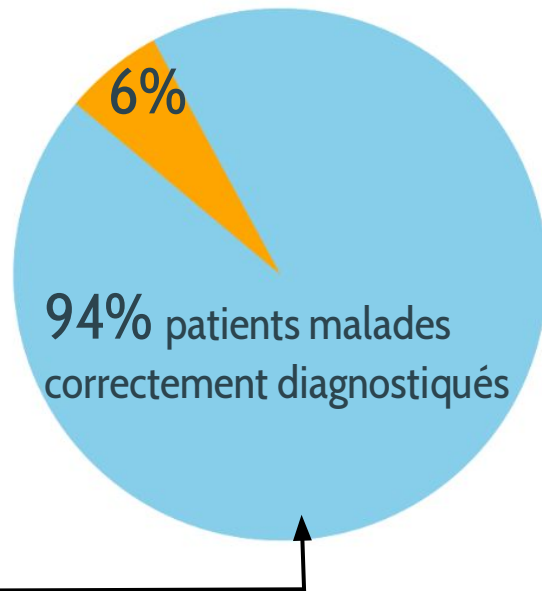


Mesure de performances : le Rappel

Prédictions

	Prédictions	
	Patients sains 	Patients malades 
Patients sains 	80	5
Patients malades 	5	80

patients malades prédits comme sains



On souhaite avoir le pourcentage le plus élevé de patients malades correctement diagnostiqués : mesure appelée le Rappel (ou Recall)



RÉSULTATS CLÉS

Performances optimales pour chaque maladie



95 % de patients correctement diagnostiqués malades



94,25 % de patients correctement diagnostiqués malades



80,65 % de patients correctement diagnostiqués malades



98 % de patients correctement diagnostiqués malades



87 % de patients correctement diagnostiqués malades





DÉPLOIEMENT DE L'OUTIL

Interface utilisateur :

- Accessible pour les professionnels de santé
- Saisie des données médicales
- Prédiction de l'état du patient
- Exportation des données





<https://innovie-web-service.onrender.com>



Données exportées à partir de l'interface utilisateur



[data_patient_diabete](#)



[data_patient_sein](#)



[data_patient_foie](#)



[data_patient_rein](#)



[data_patient_coeur](#)



*La confidentialité des données personnelles
des patients est garantie grâce à
l'anonymisation via un identifiant unique :
« ID du Patient »*





AXES D'AMÉLIORATION

- Améliorer les performances des algorithmes d'apprentissage automatique (équilibre entre la mesure du Rappel et de la Précision)
- Faciliter l'entrée des valeurs sur l'interface
- Exportation des données après prédiction et validation du diagnostic → agrandir le jeu de données de l'apprentissage automatique



Merci pour votre attention



Aloïs BRAULT
Data Analyst



Jeanne HOFFMANN
Data Analyst



Nathan GRIMMER
Data Analyst



Florence SIMONIN
Data Analyst

ANNEXE

Scores des modèles de machine learning



Modèle : RandomForestClassifier

Scaler : **QuantileTransformer**

Avg scores - Cross Validation : 0.94

Recall : 0.92

Accuracy : 0.91



Modèle : KNeighborsClassifier

Scaler : **QuantileTransformer**

Avg scores - Cross Validation : 0.9434

Recall : 0.9423

Accuracy : 0.9375



Modèle : XGBClassifier

Scaler : **MaxAbsScaler**

Avg scores - Cross Validation : 0.84

Recall : 0.85

Accuracy : 0.86



Modèle : ExtraTreeClassifier

Scaler : **QuantileTransformer**

Avg scores - Cross Validation : 0.7669

Recall : 0.8065

Accuracy : 0.8621



Modèle : GaussianNB

Scaler : **MaxAbsScaler**

Avg scores - cross Validation : 0.9705

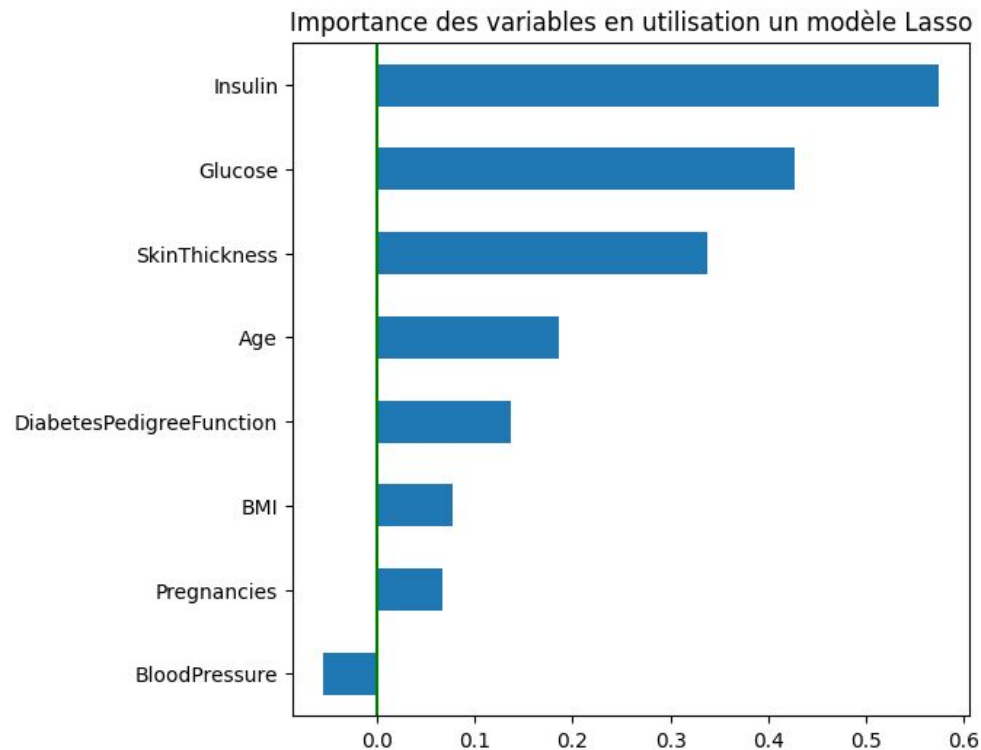
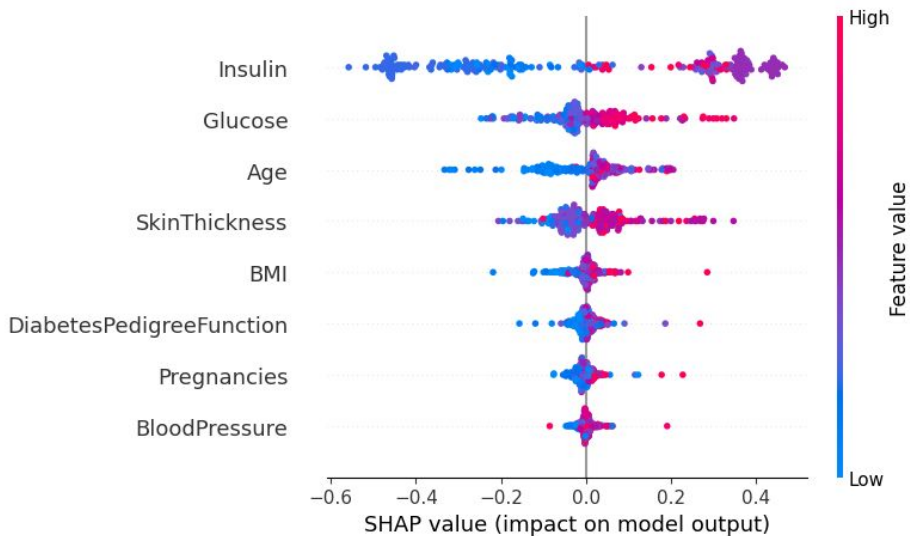
Recall : 0.9705

Accuracy : 0.9854



Diabète

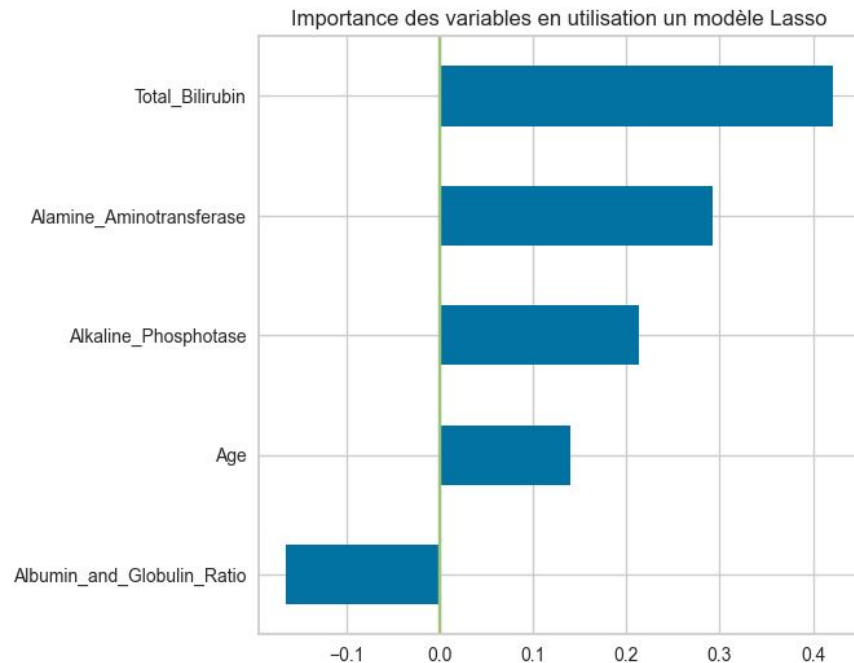
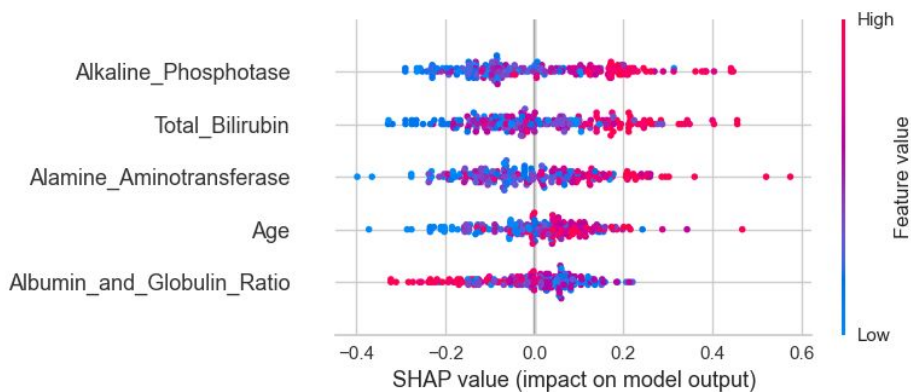
ANNEXE





ANNEXE

Maladies du foie

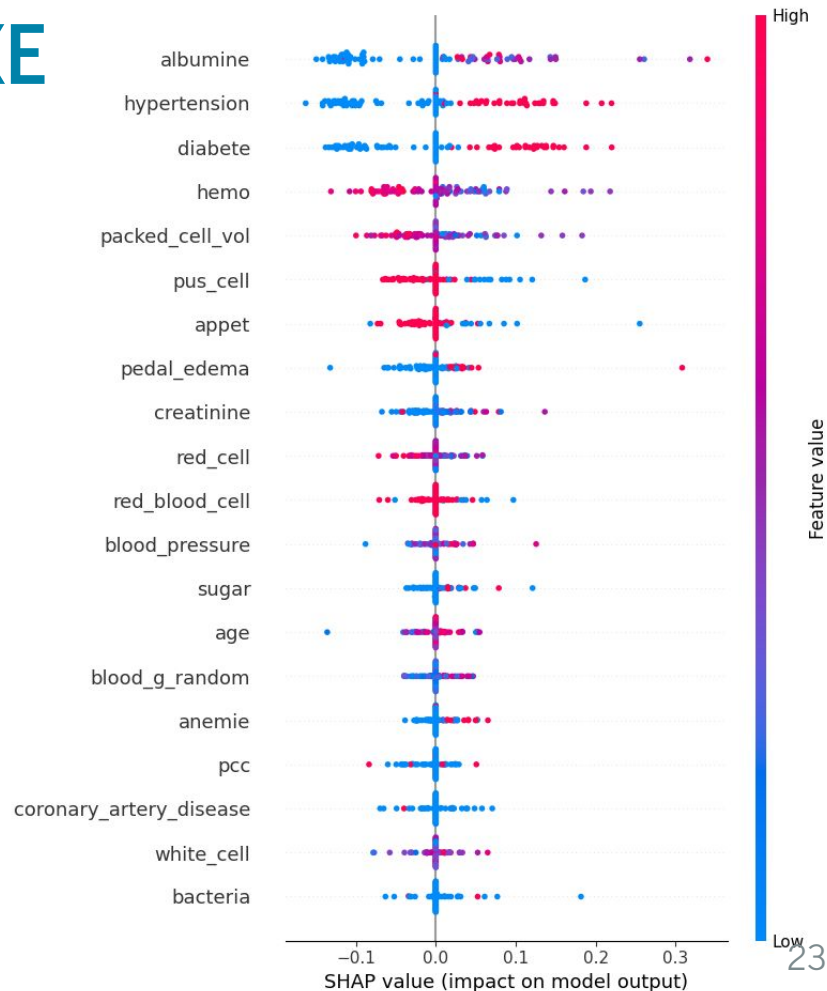
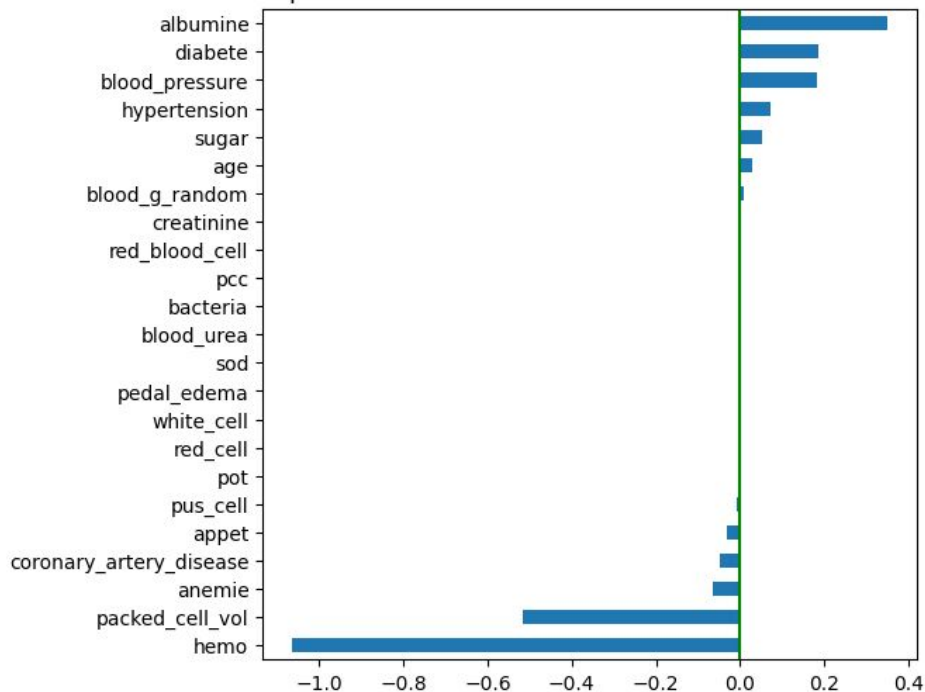




ANNEXE

Maladies rénale chroniques

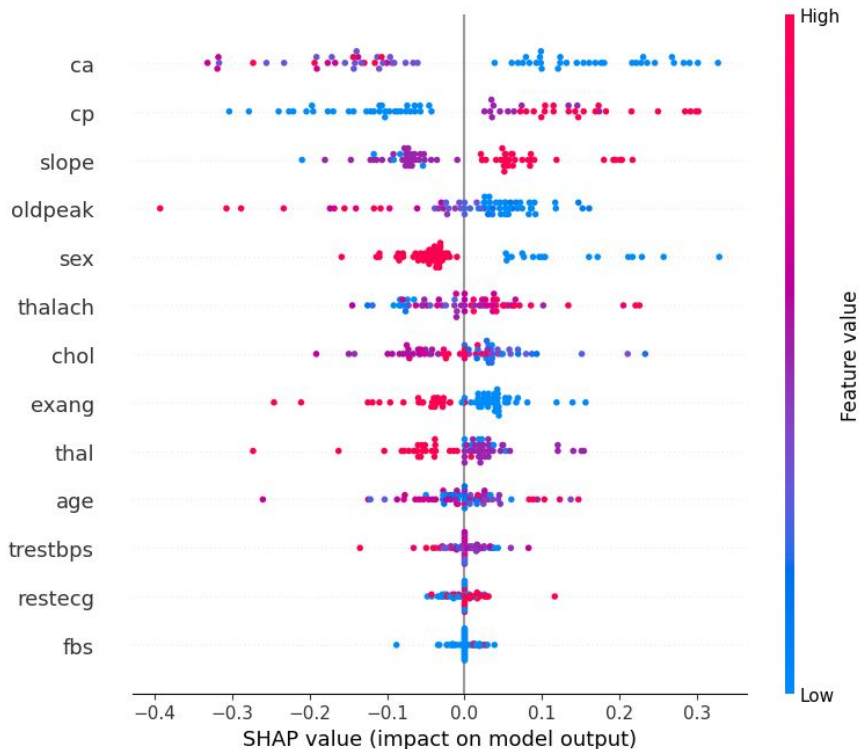
Importance des variables en utilisation un modèle Lasso



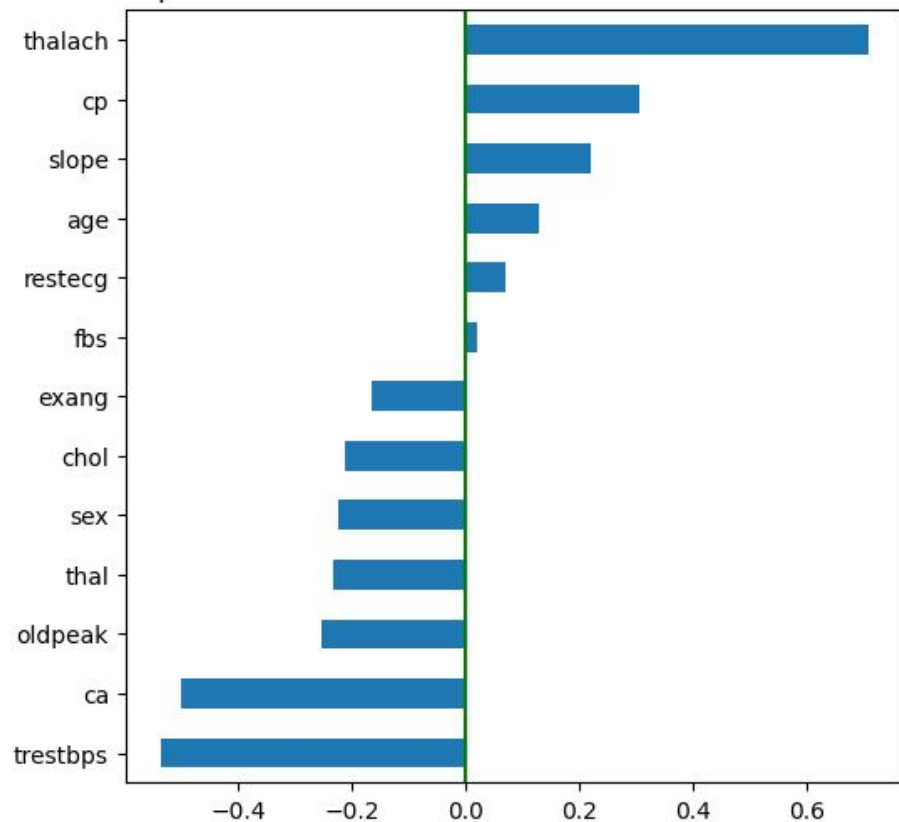


ANNEXE

Maladies cardiaques chroniques



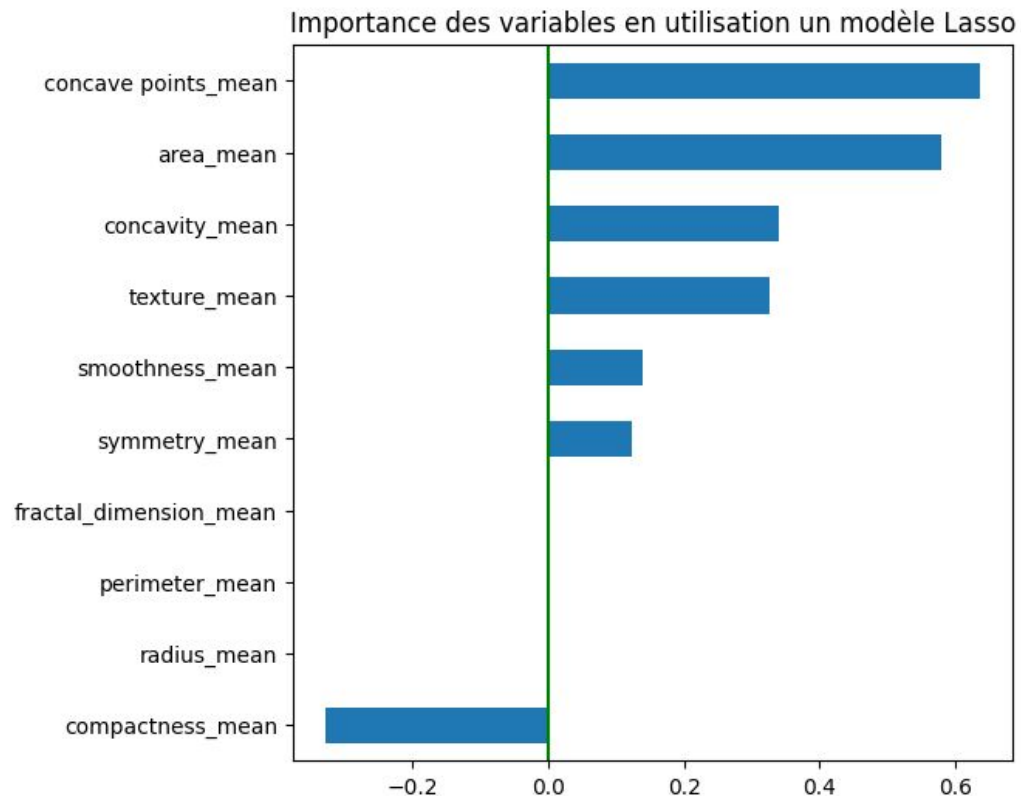
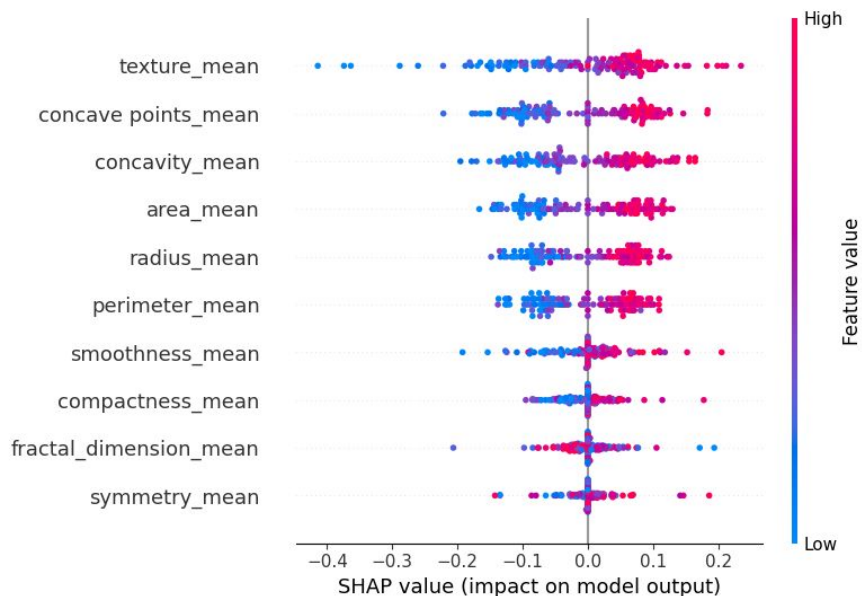
Importance des variables en utilisation un modèle Lasso





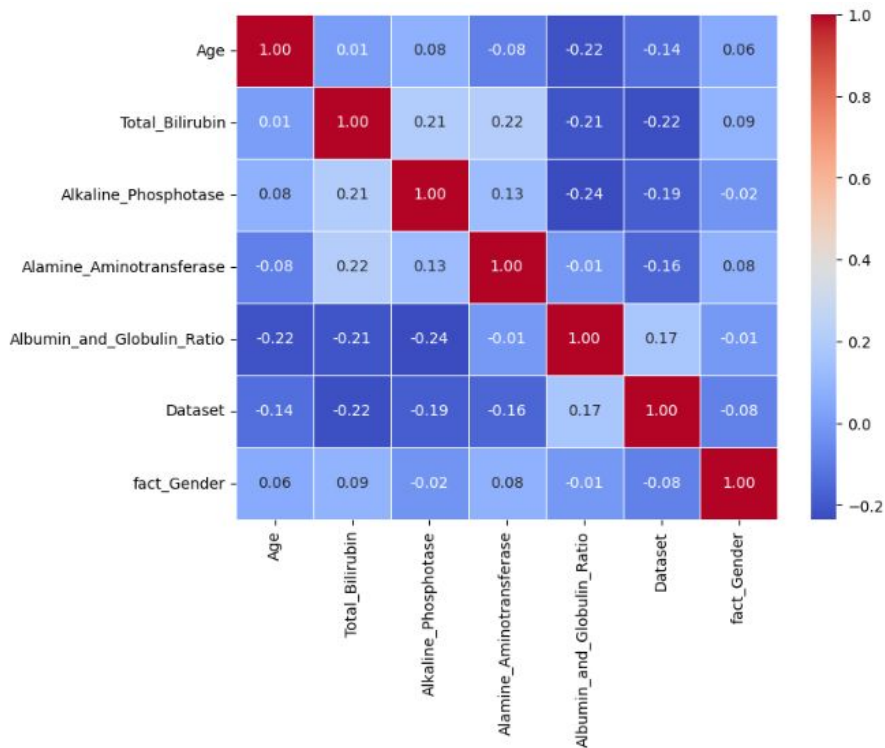
Cancer du sein

ANNEXE



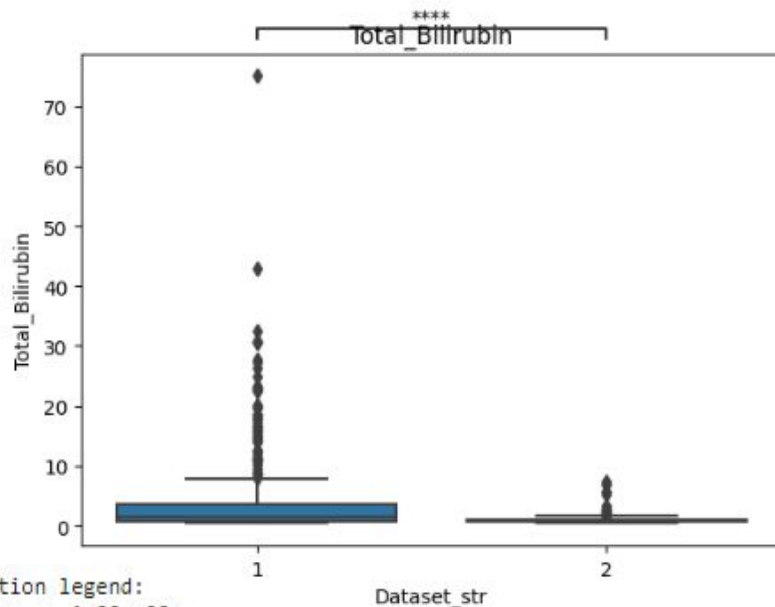
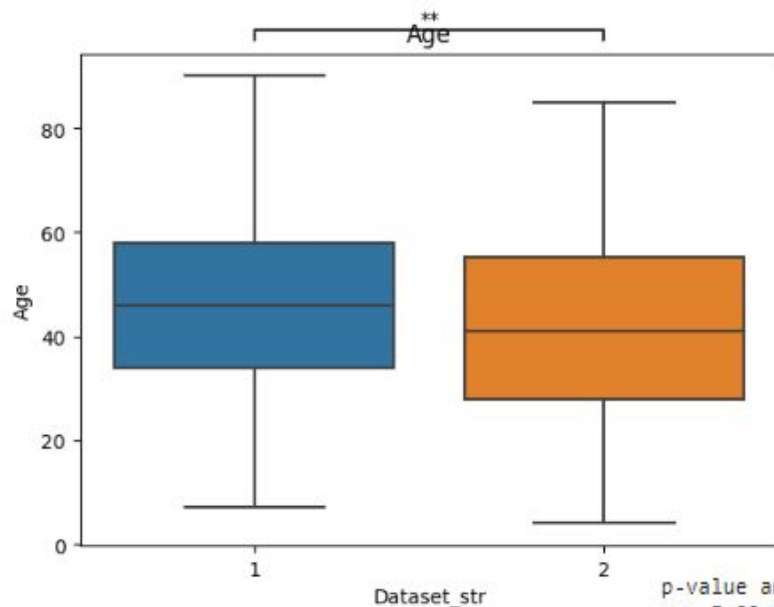
ANNEXE

Analyses : corrélations (exemple foie)



ANNEXE

Analyses : Kruskal-Wallis (différences significatives) (exemple foie)



p-value annotation legend:
ns: $5.00e-02 < p \leq 1.00e+00$
*: $1.00e-02 < p \leq 5.00e-02$
**: $1.00e-03 < p \leq 1.00e-02$
***: $1.00e-04 < p \leq 1.00e-03$
****: $p \leq 1.00e-04$