

Développement d'un système pour effectuer automatiquement un pré-diagnostic médical

Réalisé par : Mariem KALLEL
Encadré par : Pr. Yacine YADDADEN, Pr.
Yacine Benahmed et Pr. Marc-Denis
Rioux



01

Contexte



- Il y a une augmentation constante du nombre de patients due à l'apparition de plusieurs virus, notamment le COVID-19.
- Les ressources matérielles et humaines pour la prise en charge des patients sont insuffisantes.
- Cela résulte en un manque d'outils pour pré-diagnostiquer les patients afin d'éviter les cas critiques.

02

Solution proposée



Problème

- Manque de ressources humaines et matérielles.
- Système de triage des patients non optimisé.



Solution

- Créer une plateforme utilisant des algorithmes d'apprentissage automatique et d'analyse de données pour diagnostiquer les patients en se basant sur leurs signaux physiologiques et antécédents médicaux.

Objectifs du projet



1

Extraire des données à partir des bases de données MIMIC-IV et MIMIC-IV ED.



2

Entraîner des modèles pour la prédiction de l'admission et de la mortalité à l'hôpital.



3

Estimer la gravité des états des patients qui se présentent en urgence.

1- Génération de la Master Dataset

- Sélection et choix des attributs à partir de la base de données MIMIC-IV et MIMIC-IV ED (Emergency Department) relatives aux données récupérées à la station de triage.
- Obtention de **113** attributs.

2- Séparation des données par patient

- Le nombre de patients initial est supérieur à 200 milles patients.
- Pour réduire la taille des données et le temps de réponse, nous avons gardé seulement les patients ayant plus que **5** admissions en hôpital.
- Nous avons ainsi obtenu **12511** dossiers de patients.

3- Prétraitement des données

- Encodage des variables catégorielles.
- Imputation des valeurs manquantes : le KNN Imputer pour les variables catégorielles et l'imputation par moyenne pour les variables numériques.
- Normalisation des valeurs avec le MinMaxScaler().

4- Entraînement des modèles

Les deux classes (binaires) à prédire sont :

- outcome in hospital mortality
- outcome hospitalization

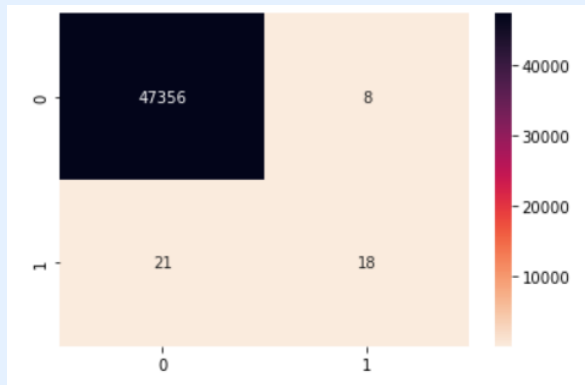
- Utilisation de 3 modèles d'apprentissage automatique : Régression Logistique (LR), le Decision Tree Classifier (DT) et le eXtreme Gradient Boosting (XGBoost).

5– Protocole de validation

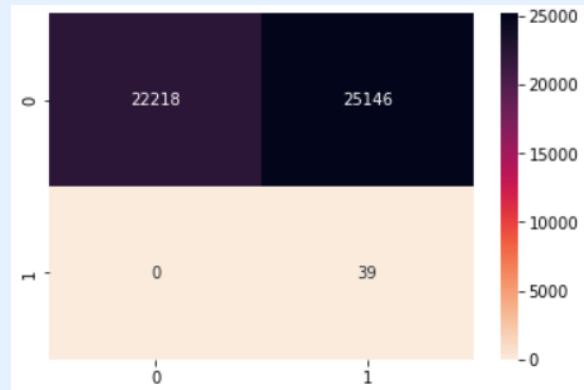
- Les tailles des données d'apprentissage et de validation sont respectivement égales à (91957, 16) et (47403, 16).
- Utilisation de la méthode de validation croisée K-fold (K=5 et K=10).

6– Evaluation des performances

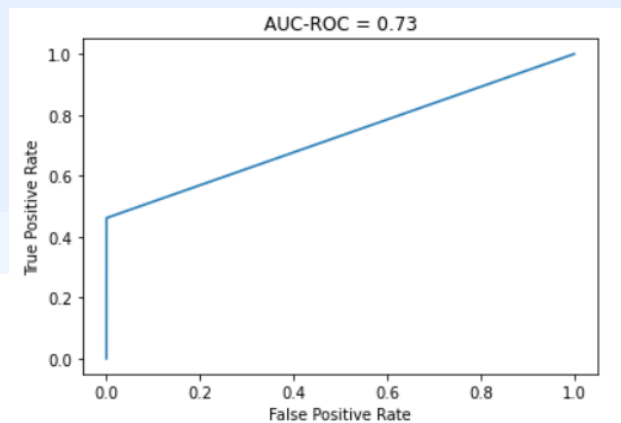
- Nous avons utilisé les métriques de performances : **l'exactitude, la précision, le rappel et la mesure F1** ainsi que **la courbe AUROC**.
- Le modèle DT a donné le meilleur résultat pour la prédiction de la classe 1 : **outcome in hospital mortality**.
- Le modèle XGB a donné le meilleur résultat pour la prédiction de la classe 2 : **outcome hospitalization**.



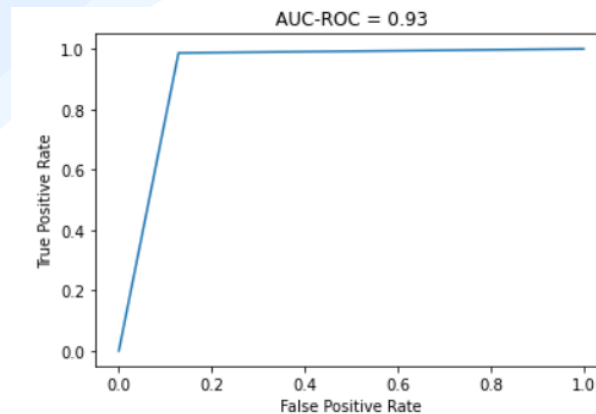
Matrice de confusion pour le modèle DT pour la classe 1



Matrice de confusion pour le modèle XGB pour la classe 2



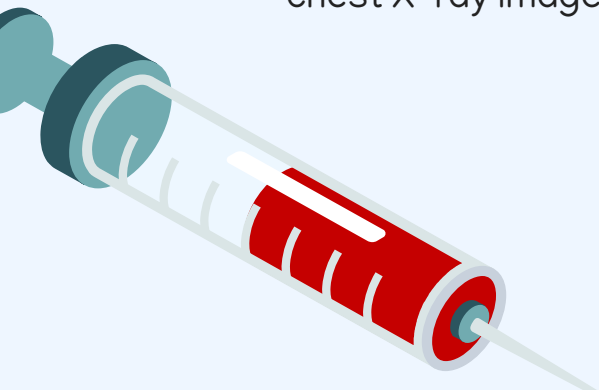
Courbe AUROC pour le modèle DT pour la classe 1



Courbe AUROC pour le modèle XGB pour la classe 2

Références

- XIE, Feng, ZHOU, Jun, LEE, Jin Wee, *et al.* Benchmarking emergency department prediction models with machine learning and public electronic health records. *Scientific Data*, 2022, vol. 9, no 1, p. 658.
- HAYAT, Nasir, GERAS, Krzysztof J., et SHAMOUT, Farah E. MedFuse: Multi-modal fusion with clinical time-series data and chest X-ray images. *arXiv preprint arXiv:2207.07027*, 2022.



**Merci pour
votre
attention!**

