



Rapport D'analyse et de prédiction :

Peut-on prédire une mort dû à un arrêt cardiaque ?

## Table des matières

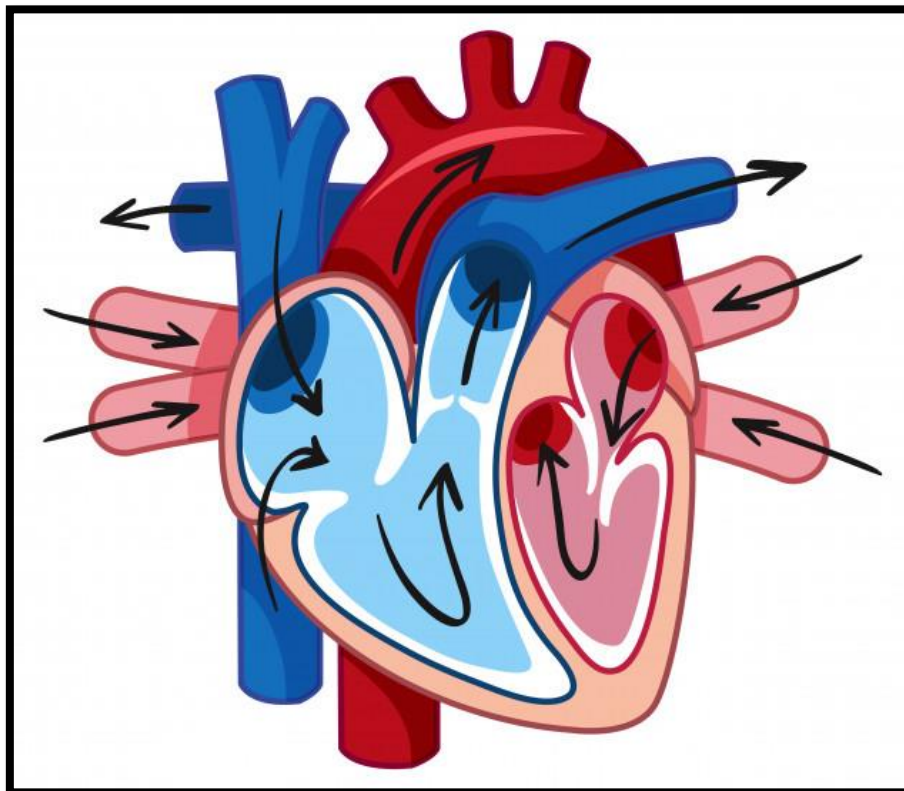
Introduction.....	2
Analyse des données .....	3

## Introduction

Via des années de progrès scientifique dans la médecine, des caractéristiques pouvant mener à un arrêt cardiaque ont été découvert. Pour aller plus loin, analyser ces caractéristiques pourraient nous permettre de comprendre les cas d'arrêt cardiaque menant à la mort. Ce document est un rapport des résultats obtenues dans le cadre de la résolution de la problématique :

### **Peut-on prédire une mort dû à un arrêt cardiaque ?**

Pour cela, nous utiliseront des données issues de patient ayant eu un arrêt cardiaque. Ces données ont été prisent à l'hôpital Allied Hospital au Faisalabad (Punjab, Pakistan) entre avril et Decembre 2015.



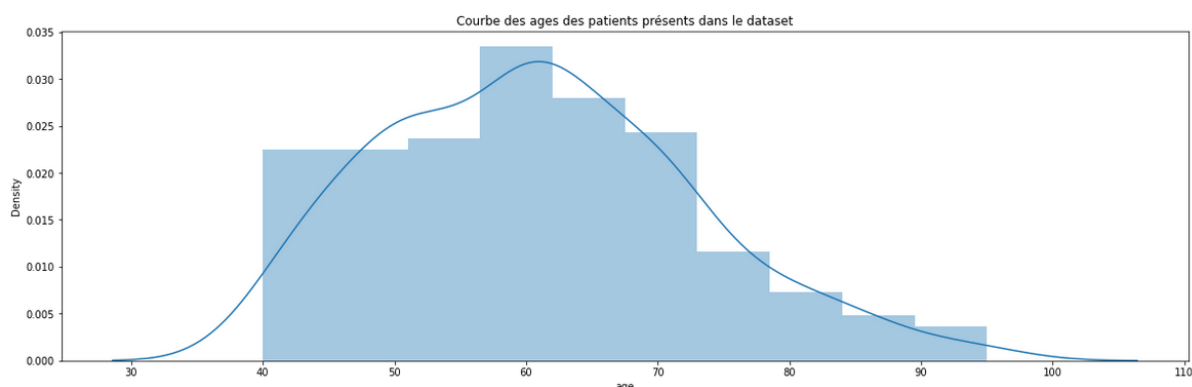
## Analyse des données

Données	Description
age	L'age du patient
anaemia	Le patient est Aménique si 1, sinon 0
creatinine_phosphokinase	Niveau de Créatine PhosphoKinase patient (mcg/L)
diabetes	Le patient est diabétique si 1, sinon 0
ejection_fraction	Fraction Ejection en pourcentage
high_blood_pressure	Le patient fait de l'hyper tension si 1, sinon 0
platelets	Nombre de plaquette du patient (kiloplatelets/mL)
serum_creatinine	Niveau de créatine dans le sang du patient (mg/dL)
serum_sodium	Niveau de sodium dans le sang du patient (mEq/L)
sex	Sexe du patient, 1 pour homme et 0 pour femme
smoking	Le patient est un fumeur si 1, sinon 0
time	Nombre de jours passé après de commencement des enregistrement
DEATH_EVENT	Le patient est mort si 1, sinon 0
Day	Jour de la mesure
Month	Mois de la mesure
Year	Année de la mesure
Date	Date de la mesure

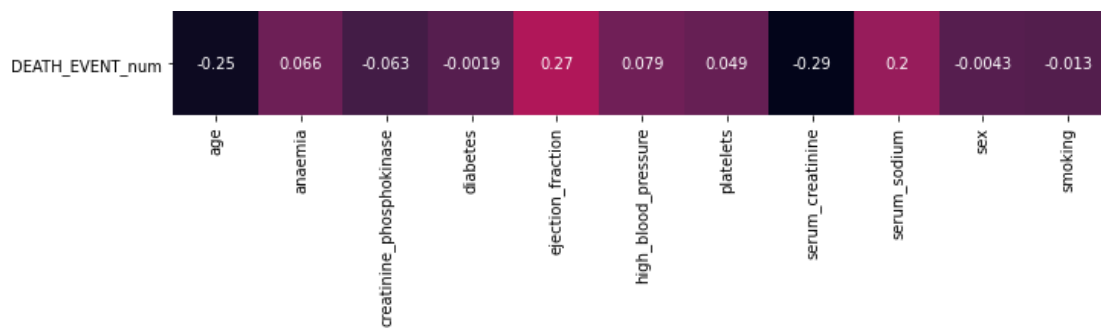
L'Anémie est le manque d'hémoglobine. C'est manque qui entraîne un mauvais transport par le sang du dioxygène et cause notamment une pâleur de la peau et des muqueuses.

La Créatine PhosphoKinase est une protéine importante dans le métabolisme énergétique. Elle est connue pour être en relation avec le diagnostic d'un infarctus.

La fraction d'éjection est le pourcentage d'éjection du sang contenu dans une cavité cardiaque lors d'un battement.



Le dataset proposé est contient des données de patient entre 40 et 95 ans. La prédiction ne sera donc valable que pour cette tranche d'âge-là.

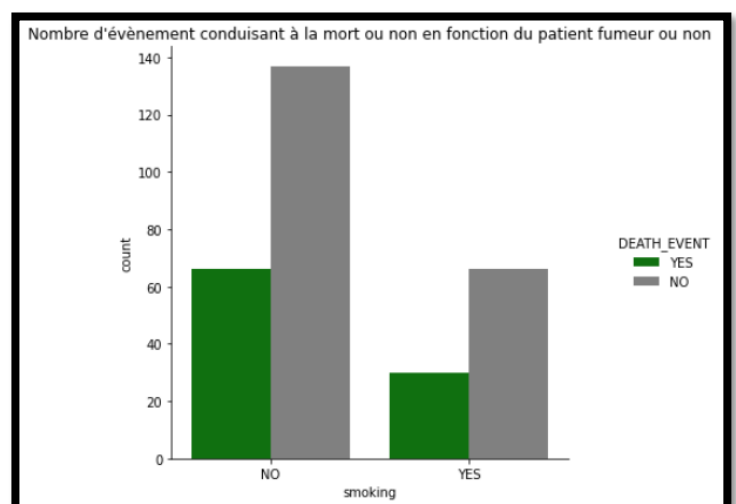
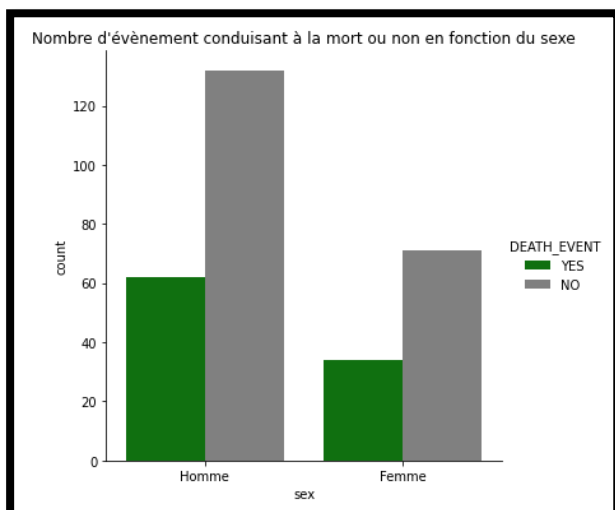
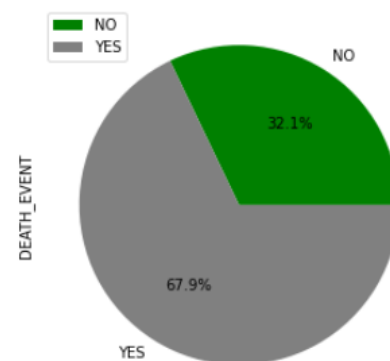


Nous avons remarqué que dans le dataset fourni, les données proposées pour prédire la mort du patient ne présentent aucune corrélation forte avec cette dernière. Nous avons néanmoins continué notre analyse pour repérer les différentes éventualités.

On remarque aussi que le nombre de personnes mortes présentées dans ce dataset est plus faible que le nombre de personnes ayant survécu à l'arrêt cardiaque.

L'étude rapide durant ces 2 jours ne nous a pas permis d'extraire des features plus importantes que d'autres. Nous avons donc utilisé les données jugées utiles par les médecins. Voici un exemple des différents graphes où l'on peut voir que les différences entre la mort ou non d'un arrêt cardiaque ne sont pas forcément fortement liées à une donnée.

Diagramme circulaire du nombre de mort de ce dataset (en pourcentage)



## Résultat

Après avoir étudié les données, nous avons entraînés plusieurs algorithmes pour prédire la mort ou non du patient ayant l'arrêt cardiaque. Nous avons ensuite testé ces algorithmes avec des données de test et voici la liste des algorithmes entraînés et leurs scores de prédictions :

<b>LinearSVC</b>	• 74,8%
<b>RandomForest</b>	• 78,0%
<b>DecisionTree</b>	• 74,6%
<b>LogisticRegression</b>	• 74,8%

On remarque que l'algorithme ayant un meilleur score de prédiction est le RandomForest. C'est cet algorithme qui a été livré. Cet algorithme reste néanmoins perfectible. Pour cela, nous avons réalisé un container docker permettant de stocker les données, de les récolter et de prédire grâce à ces données récolté les futurs cas.

Docker		
Base de données PostgreSQL (127.0.0.3:5432)	API FastAPI (172.0.0.2:8001)	Site Web Streamlit (172.0.0.4:8501)

## Conclusion

Malheureusement, les données fournies ne permettent pas d'avoir une certitude sur la possibilité du patient à survivre à un arrêt cardiaque. Le manque de données et de temps d'analyse en est la cause.