

Le projet que j'ai choisi est celui de l'[UCI Heart Disease Data](#)

### **Type de problème et enjeux du projet**

La tâche principale du projet est d'identifier les personnes à risque de développer des maladies du cœur : donc il est possible de parler d'un type de problème **médical**.

L'enjeu du projet est d'identifier des tendances et des patterns permettant aux médecins de pouvoir déceler plus rapidement chez leurs patients à risque ce type de problème.

### **Analyse du contexte**

La base de données est multivariée et composée de 14 attributs. À ce jour, elle est la seule à avoir été utilisée par des chercheurs utilisant l'apprentissage machine.

La tâche principale de ce projet est d'identifier les personnes à risque de développer des maladies du cœur. Sans oublier la tâche expérimentale de diagnostiquer et de découvrir d'autres informations qui pourraient aider à mieux comprendre le problème.

### **Définition des objectifs et hypothèses**

L'un des objectifs principaux selon moi est d'identifier les facteurs à prendre en compte parmi les 14 attributs donnés. Dans le but de pouvoir dresser dans les prochaines étapes un modèle fiable.

Rapidement sans avoir consulté et assimilé tous les paramètres : mon hypothèse selon mon bagage personnel est que les personnes plus âgées, ayant un taux de cholestérol élevé sont plus à risque de développer des maladies cardiovasculaires.

### **Description du sommaire de la méthodologie**

En transformant les données via MongoDB, je m'assurerai de conserver uniquement les données pertinentes.

Par la suite, j'analyserai ce problème avec Spark MLlib : un modèle de classification sera créé avec les paramètres jugés pertinents.