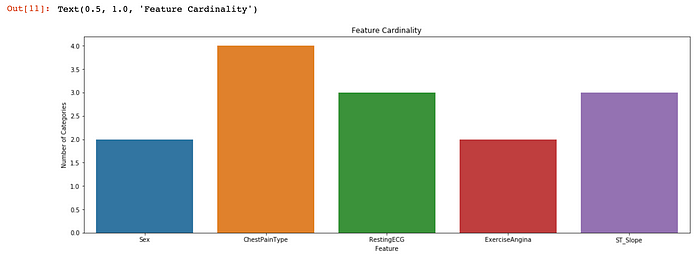
**Introduction générale**

Les maladies cardiovasculaires sont l'une des principales causes de décès dans le monde. Par conséquent, si l'apprentissage automatique pouvait aider à prédire la prédiction de l'insuffisance cardiaque, la contribution serait significative. L'ensemble de données utilisé dans ce didacticiel a été développé par Davide Chicco, Giuseppe Jurman de BMC Medical Informatics and Decision Making. Il a été open-source et peut être téléchargé à partir de [Kaggle](https://www.kaggle.com/fedesoriano/heart-failure-prediction) .

Les données contiennent un total de 918 instances (ou lignes) avec 12 attributs (ou colonnes). Sur ces 12 attributs, 5 sont de nature catégorielle et 7 sont de nature numérique. Continuons maintenant et importons les bibliothèques nécessaires.

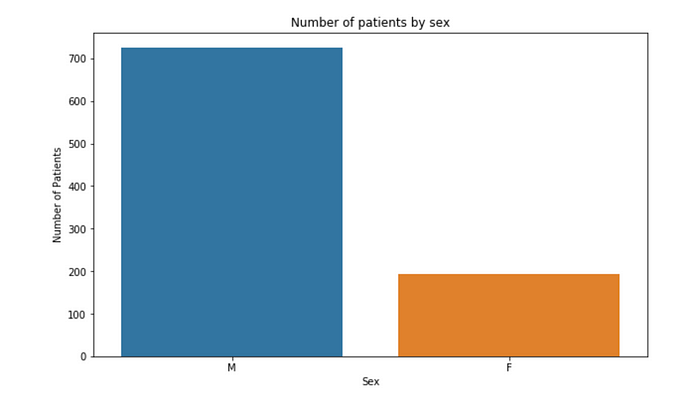
**Introduction**

la prédiction de l'insuffisance cardiaque est un problème important dans le domaine médical, et les algorithmes d'apprentissage automatique peuvent être utilisés pour développer des modèles prédictifs précis à l'aide d'ensembles de données accessibles au public. Plusieurs algorithmes peuvent être utilisés, et le choix de l'algorithme dépend des caractéristiques de l'ensemble de données et du problème à résoudre.

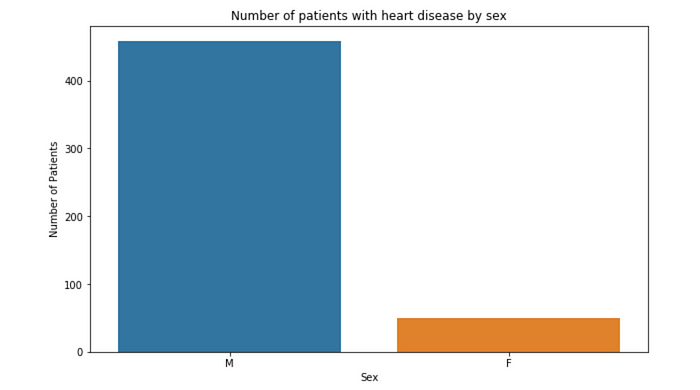


Cardinalité de chaque caractéristique catégorielle

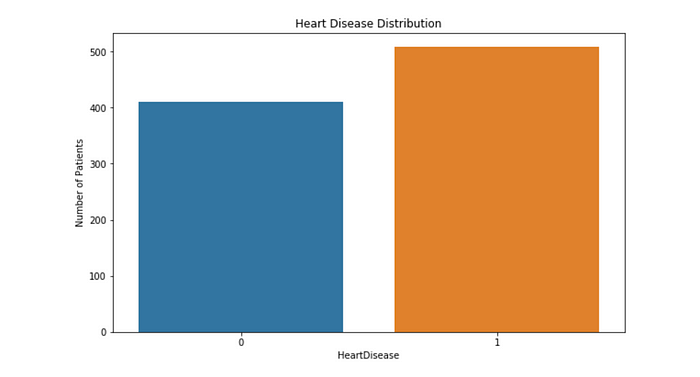
Toutes les caractéristiques catégorielles ont une cardinalité relativement faible, ce qui facilite l'encodage et le traitement de ces caractéristiques par le modèle. Une propriété intéressante de cet ensemble de données est qu'environ 79 % des patients étaient des hommes. De plus, plus de 90 % des patients atteints de maladies cardiaques étaient des hommes.



Beaucoup plus de patients étudiés étaient des hommes



Plus de 90 % des patients atteints de maladies cardiaques étaient des hommes

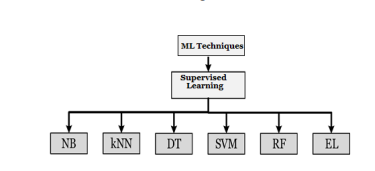


Répartition des maladies cardiaques

**Arrière-plan**

**Techniques d'apprentissage automatique (ML) pour la classification**

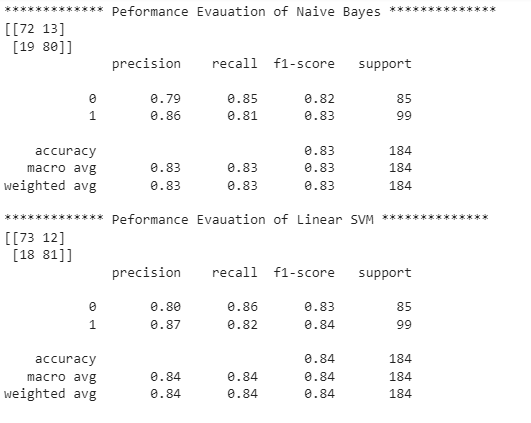
Ici, diverses méthodes d'analyse des classificateurs basés sur l'apprentissage automatique. Ceci est montré dans Figure 1



Techniques de ML pour la classification

**Classification**

Pour comparer les performances de différents algorithmes d'apprentissage automatique pour la prédiction de l'insuffisance cardiaque, nous pouvons considérer Naive Bayes, Support Vector Machines (SVM) et Neural Networks. Ce sont tous des algorithmes populaires pour les problèmes de classification, et chacun a ses propres forces et faiblesses.



**Naive Bayes** est un algorithme simple basé sur le théorème de Bayes. Il est particulièrement utile lorsqu'il s'agit d'ensembles de données de grande dimension, car il est efficace en termes de calcul et peut gérer un grand nombre d'entités. Naive Bayes suppose que les caractéristiques sont conditionnellement indépendantes compte tenu de la variable cible, ce qui n'est pas toujours vrai dans la pratique. Malgré cette hypothèse simplificatrice, Naive Bayes peut souvent atteindre une grande précision dans la pratique.

**Les machines à vecteurs de support (SVM)** sont un autre algorithme populaire pour les problèmes de classification. Ils fonctionnent en trouvant l'hyperplan dans un espace de grande dimension qui sépare au maximum les différentes classes. Les SVM peuvent gérer des données non linéairement séparables en mappant les caractéristiques d'entrée à un espace de dimension supérieure. Cependant, les SVM peuvent être coûteuses en calcul pour les grands ensembles de données, et choisir la bonne fonction de noyau peut être difficile.

**Les réseaux de neurones** sont une classe puissante d'algorithmes d'apprentissage automatique inspirés de la structure et de la fonction du cerveau. Les réseaux de neurones peuvent apprendre des relations complexes et non linéaires entre les caractéristiques d'entrée et la variable cible, et peuvent souvent atteindre des performances de pointe sur un large éventail de problèmes de classification. Cependant, les réseaux de neurones peuvent être coûteux en calcul à former et peuvent nécessiter de grandes quantités de données pour bien se généraliser.

**En résumé**, Naive Bayes, SVM et Neural Networks sont tous des algorithmes populaires pour la prédiction de l'insuffisance cardiaque, et chacun a ses propres forces et faiblesses. Naive Bayes est efficace en termes de calcul et peut gérer des ensembles de données de grande dimension, mais fait l'hypothèse simplificatrice de l'indépendance des fonctionnalités. Les SVM peuvent gérer des données non linéairement séparables, mais peuvent être coûteuses en calcul et nécessitent un choix judicieux de la fonction du noyau. Les réseaux de neurones sont puissants et flexibles, mais peuvent être coûteux en calcul et nécessiter de grandes quantités de données pour bien généraliser. Le choix de l'algorithme dépendra des caractéristiques de l'ensemble de données et du problème spécifique à résoudre.