

**Année universitaire : 2021/2022**

***Réalisé par*** : ***Encadré par :***

ENNANI Abdelmalek Mr. QBADOU Mohamed

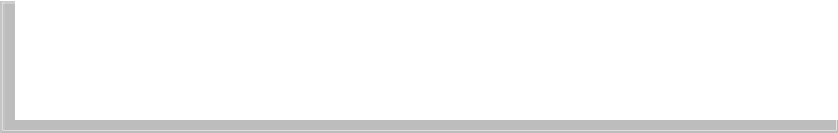
DOUHI Chaimae

**Rapport :**

**Filière :**

**‘ Ingénierie Informatique : Génie du logiciel et des Système Informatiques Distribués ’**

**Département Mathématique et Informatique**



Réalisation d’un projet

Prédiction d'insuffisance cardiaque utilisant l'algorithme KNN

**Sommaire :**

Introduction ------------------------------------------------------------------- 3

Contexte Général du projet ------------------------------------------------- 4

L’algorithme KNN ------------------------------------------------------------- 5

Étapes de fonction de KNN--------------------------------------------------- 6

Data-set ------------------------------------------------------------------------- 7

Format de data-set ------------------------------------------------------------ 8

Thème de projet --------------------------------------------------------------- 9

Outils utilisés ------------------------------------------------------------------- 9

Les composants de programme -------------------------------------------- 10

Informations requises par le programme --------------------------------- 11

Le code : ------------------------------------------------------------------------------ 12

Disease.h ----------------------------------------------------------- ----------------- 12  
Knn-funcs**.**c ------------------------------------------------------------------------- 13  
Main.C ------------------------------------------------------------------------------- 19

Exécution de code ------------------------------------------------------------------ 20

Conclusion ---------------------------------- ----------------------------------------- 21

**Introduction**

Les algorithmes sont bien plus présents dans nos vies qu’on ne le pense. En effet, ces calculs informatiques sont la base de la majorité des logiciels, de l’intelligence artificielle et de la plupart de nos programmes et services. Aujourd’hui, ces calculs d’entiers et les structures de données engendrées sont la base et l’essence même de notre société.

De nos jours, écrire un algorithme se fait désormais sur ordinateur, pour effectuer une tâche ou un programme informatique. Or, pour être fonctionnel, il est nécessaire « d’indiquer » à l’ordinateur ce qu’il est censé faire et ce, avec précision et étape par étape.

L'un parmi ces algorithmes est l'algorithme KNN.

**Contexte Général du projet :**

Les maladies cardiovasculaires (MCV) sont la première cause de décès dans le monde, prenant environ 17,9 millions de vies chaque année, ce qui représente 31 % de tous les décès dans le monde. Quatre décès par 5CVD sont dus à des crises cardiaques et à des accidents vasculaires cérébraux, et un tiers de ces décès surviennent prématurément chez des personnes de moins de 70 ans. L'insuffisance cardiaque est un événement courant causé par les maladies cardiovasculaires et cet ensemble de données contient 11 caractéristiques qui peuvent être utilisées pour prédire une éventuelle maladie cardiaque.

Les personnes atteintes de maladies cardiovasculaires ou à haut risque cardiovasculaire (en raison de la présence d'un ou plusieurs facteurs de risque tels que l'hypertension, le diabète, l'hyperlipidémie ou une maladie déjà établie) ont besoin d'une détection et d'une prise en charge précoces dans lesquelles un modèle d'apprentissage automatique peut être d'une grande aide.

C'est pourquoi nous avons décidé de développer un programme utilisant le langage C, un programme qui nous permet de prédire l'insuffisance cardiaque sur la base d'un ensemble de données qui contient beaucoup d'informations sur cette maladie. Pour cela, nous avons utilisé l'algorithme

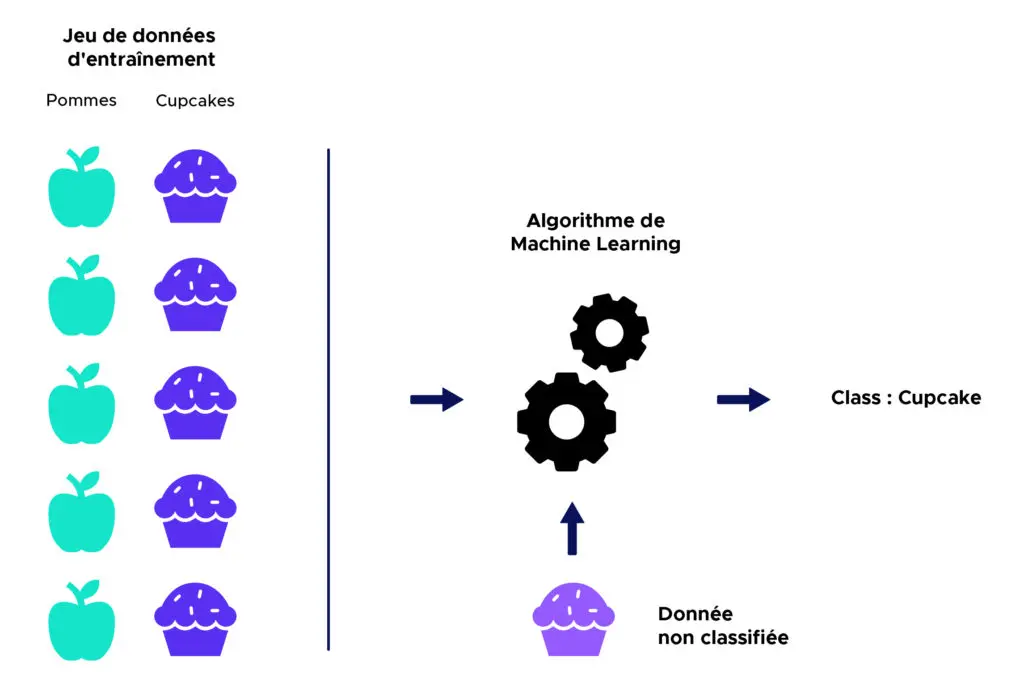
## KNN (K plus proches voisins ou K-nearest neighbors)

# L’algorithme KNN :



L’algorithme des K plus proches voisins ou K-nearest neighbors (kNN) est un algorithme de Machine Learning qui appartient à la classe des algorithmes d’apprentissage supervisé simple et facile à mettre en œuvre qui peut être utilisé pour résoudre les problèmes de classification et de régression.

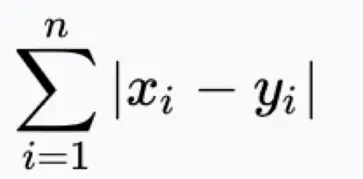
Cette algorithme reçoit un ensemble de données qui est étiqueté avec des valeurs de sorties correspondantes sur lequel il va pouvoir s’entraîner et définir un modèle de prédiction. Cet algorithme pourra par la suite être utilisé sur de nouvelles données afin de prédire leurs valeurs de sorties correspondantes.

Ex : 

L'intuition derrière l'algorithme des K plus proches voisins est l’une des plus simples de tous les algorithmes supervisés :

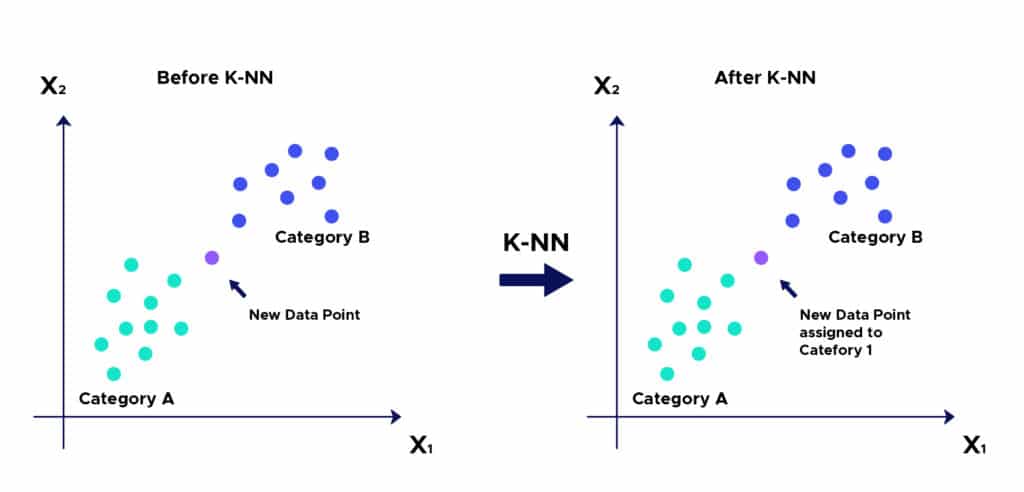
* **Étape 1** : Sélectionnez le nombre K de voisins
* **Étape 2**: Calculez la distance Du point non classifié aux autres points.

On utilise l’EUCLIDIENNE pour calculer la distance :



* **Étape 3**: Prenez les K voisins les plus proches selon la distance calculée.
* **Étape 4** : Parmi ces K voisins, comptez le nombre de points  appartenant à chaque catégorie.
* **Étape 5**: Attribuez le nouveau point à la catégorie les plus présents parmi ces K voisins.

***Exemple*** :



Data-set Utilisé :

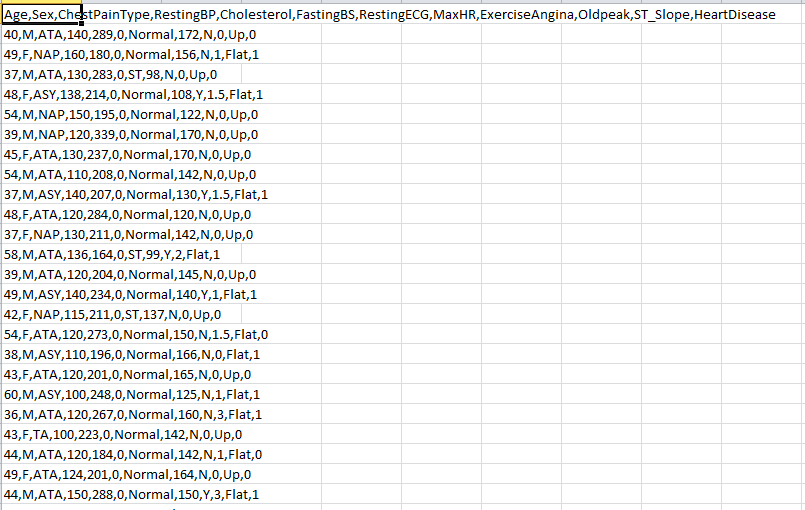
Ce jeu de données a été créé en combinant différents jeux de données déjà disponibles indépendamment mais non combinés auparavant. Dans cet ensemble de données, 5 ensembles de données cardiaques sont combinés sur 11 caractéristiques communes, ce qui en fait le plus grand ensemble de données sur les maladies cardiaques disponible à ce jour à des fins de recherche. Les cinq ensembles de données utilisés pour sa conservation sont :

* Cleveland : 303 observations
* Hongrois : 294 observations
* Suisse : 123 observations
* Long Beach Virginie : 200 observations
* Ensemble de données Stalog (cœur): 270 observations
* Total : 1190 observations
* Dupliqué : 272 observations

*Informations sur les attributs :*

1. Âge : âge du patient [années]
2. Sexe : sexe du patient [H : Masculin, F : Féminin]
3. ChestPainType : type de douleur thoracique [TA : Angine typique, ATA : Angine atypique, NAP : Douleur non angineuse, IS : Asymptomatique]
4. RestingBP: tension artérielle au repos [mm Hg]
5. Cholestérol : cholestérol sérique [mm/dl]
6. FastingBS : glycémie à jeun [1 : si FastingBS > 120 mg/dl, 0 : sinon]
7. RestingECG: résultats de l'électrocardiogramme au repos [Normal : normal, ST : présentant une anomalie de l'onde ST-T (inversions de l'onde T et/ou élévation ou dépression du segment ST > 0,05 mV), HVG : montrant une hypertrophie ventriculaire gauche probable ou certaine selon les critères d'Estes]
8. MaxHR: fréquence cardiaque maximale atteinte [Valeur numérique entre 60 et 202]
9. ExerciseAngina: exercise-induced angina [Y: Oui, N: Non]
10. Oldpeak: oldpeak = ST Valeur numérique mesurée en dépression]
11. ST\_Slope: the slope of the peak exercise ST segment [Up: upsloping, Flat: flat, Down: downsloping]
12. HeartDisease: output class [1: heart disease, 0: Normal]

Format de data-set :

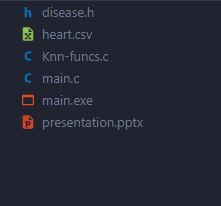


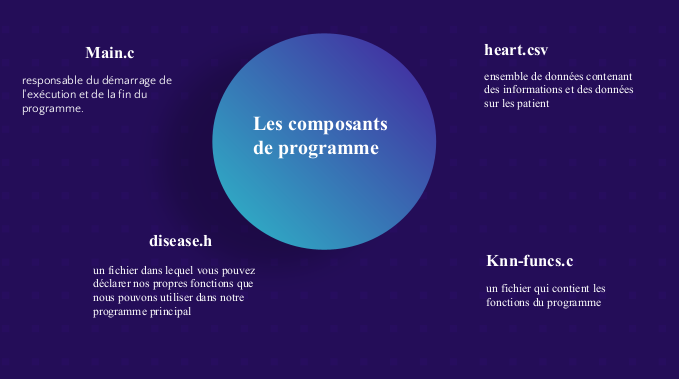
**Thème de projet :**Un projet programmé dans le langage de programmation C qui nous permet de savoir si une personne a une maladie cardiaque ou non, en utilisant l'algorithme kNN et le DATASET qui contiennent des informations de personnes atteintes de maladie cardiaque et de personnes non cardiaques

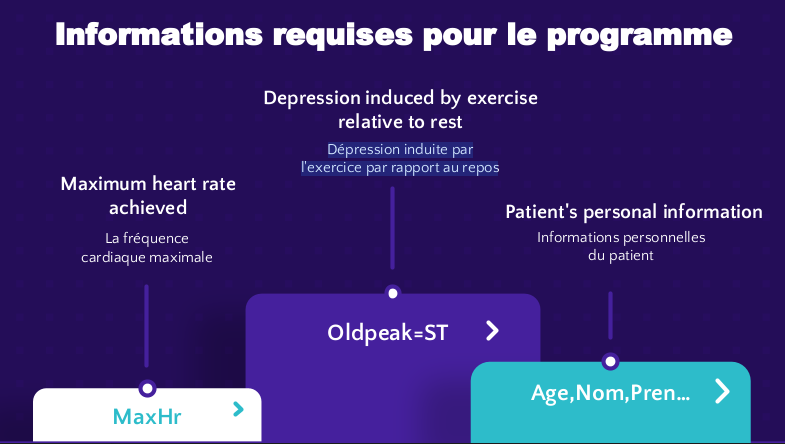
**Outils Utilisé :**

* **Langage C :** est un langage de programmation informatique procédural à usage général prenant en charge la programmation structurée, la portée des variables lexicales et la récursivité, avec un système de type statique. De par sa conception, C fournit des constructions qui correspondent efficacement aux instructions machine typiques
* **Heart.csv :** data-set qui contient les informations sur la maladie cardiaque
* **Vs code :** Visual studio code ou VS Code est un éditeur de code développé par Microsoft en 2015. Contrairement à ce à quoi Microsoft a eu l’habitude de nous habituer durant des années, il est l’un de ces premiers produits open source et gratuit, et surtout disponible sur les systèmes d’exploitation Windows, Linux et Mac.

**Fichiers de programme:**

****





**disease.h :**

****

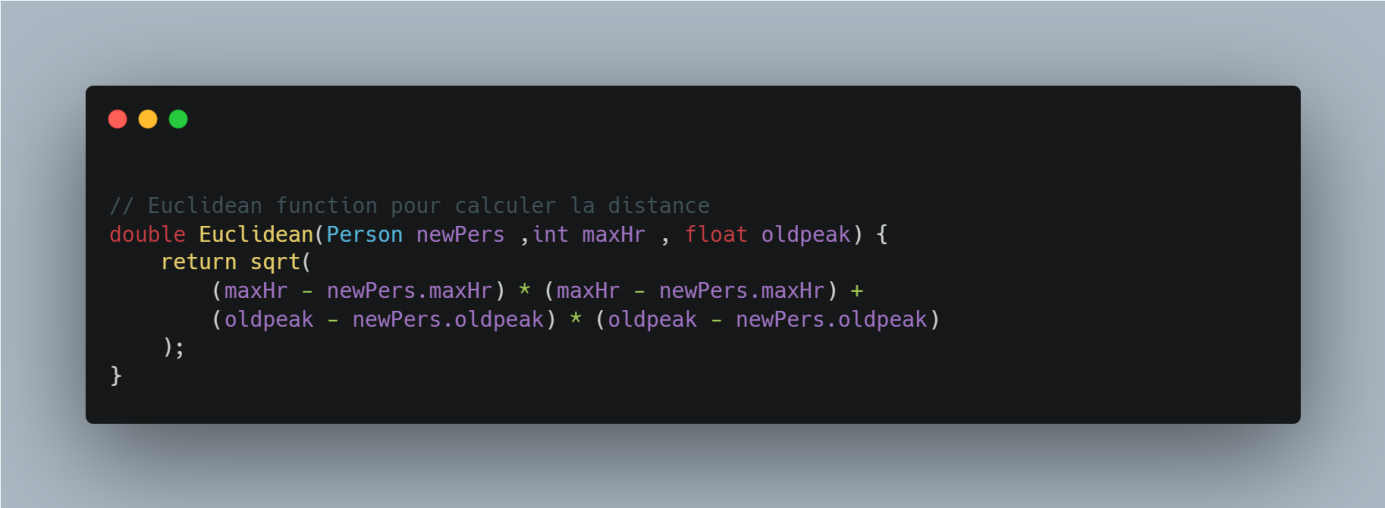
**Knn-funcs.c :**

****Pour pinter des informations sur le programme (header informations)

  
Une fonction pour saisir les informations de la personne dont nous souhaitons vérifier si elle souffre ou non d'une maladie cardiaque. La fonction demandera le nom, le prénom, l'âge et maxHer ainsi que Oldpeak de la personne

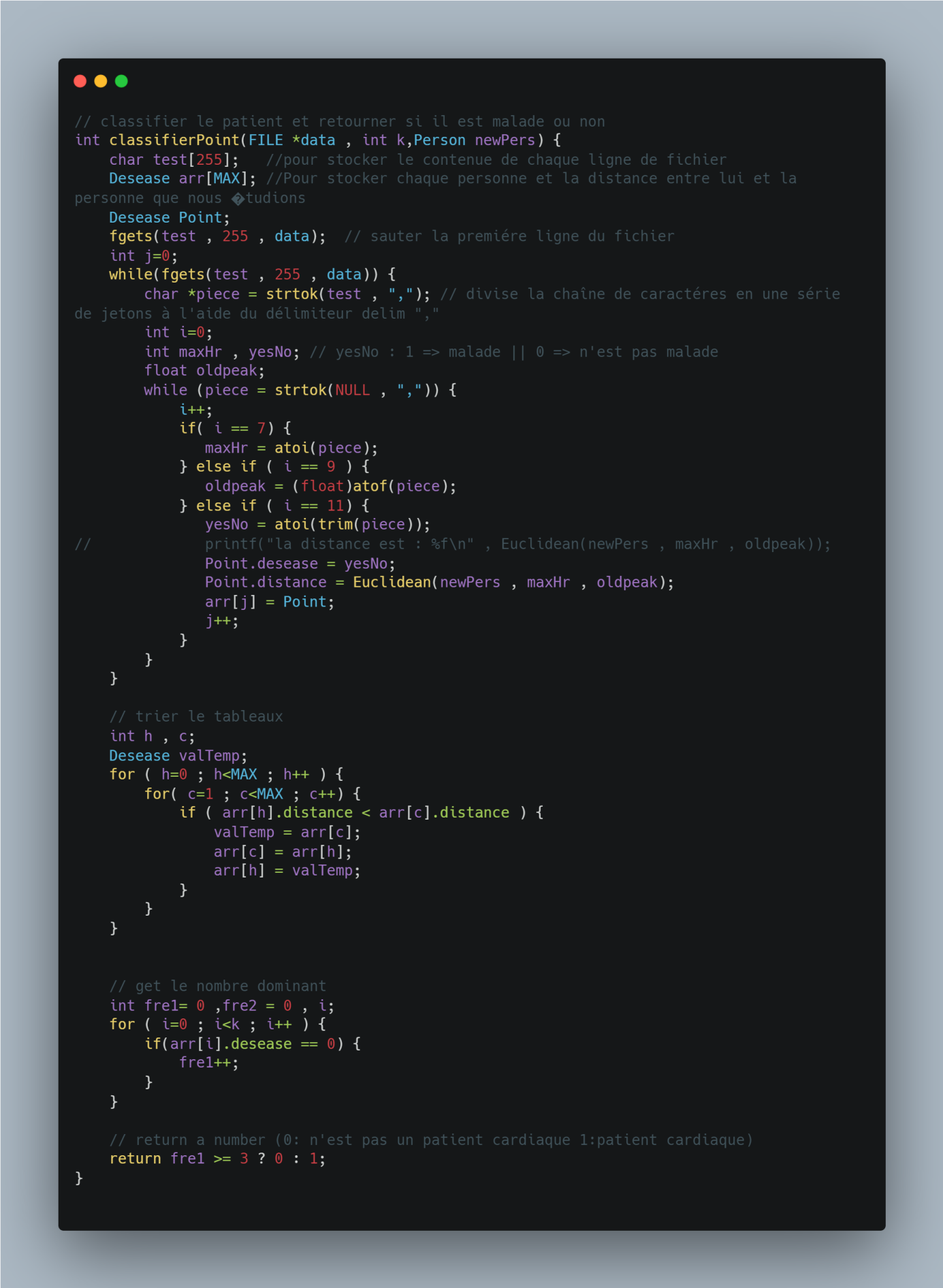


La fonction printPerson vous permet de printer les informations du patient

  
La fonction Euclidean vous permet de calculer la distance euclidienne entre la personne saisie et chaque personne dans l'ensemble de données



Les fonctions trim, ltrim et rtrim permettent de supprimer les espaces avant et après une chaîne



La fonction classifierPoint permet de :

* Recevoir le fichier, la valeur k et les informations personnelles comme

parametres

* lire chaque ligne de la fichier heart.csv
* extraire les valeurs nécessaire de chaque ligne (max et oldPeak)
* Mémorisation de la valeur de la maladie (0 ou 1) et de la distance euclidienne entre chaque personne du fichier et la personne dont on veut savoir si il estmalade ou non dans un tableau
* Trier le tableau
* Retourner le nombre dominant dans le premier K elements dans le tableau
* Si 1 est le nombre dominant donc la personne est malade
* Si 0 est le nombre dominant donc la personne n’est pas malade

**Main.C fichier :**

****Le fichier main.c permet d’executer notre programme de langage C.

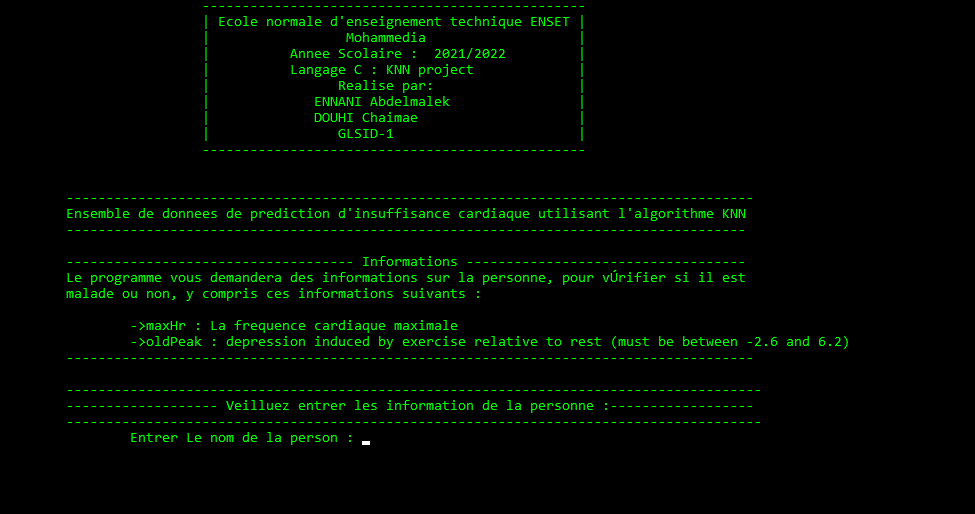
Il contient des variable nécessaire pour notre programme, comme la variable data qui a le typeFILE , cette variable permet d’ouvrir un fichier à l’aide de la fonction fopen.

- la variable newPerson pour stocker les information de la personne

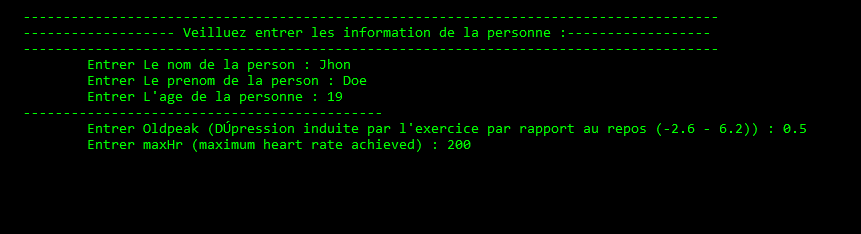
- la variable k pour les k premiers éléments et seulement dans le tableau

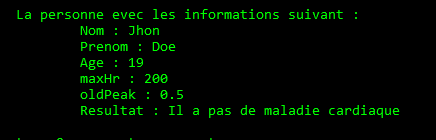
- la variale res pour stocker la valeur retourné par classifierPoint fonction

**Comment fonctionne le programme**:



Le programme affichera des informations relatives au programme et à ses objectifs, puis il demandera des informations importantes pour que le programme fonctionne, telles que le nom, le prenom, l'âge et La fréquence cardiaque maximale, ansi que Dépression induite par l'exercice par rapport au repos (valeur entre 0.2 et 0.6)





Après avoir entré toutes les informations requises, le programme prendra ces informations et classera la personne selon qu'elle appartient au groupe malade ou au groupe sain et en conséquence, il nous dira si la personne est malade ou non



si l'utilisateur souhaite entrer les informations d'une autre personne il lui suffit de taper 0 et d'appuyer sur entrée , le programme se rechargera et lui donnera la possibilité d'entrer les informations d'une autre personne**Conclusion :**

Data-set : [Heart Failure Prediction Dataset | Kaggle](https://www.kaggle.com/fedesoriano/heart-failure-prediction)

Notre Mini Projet a pour but la mise en place une programme de prédiction de l'insuffisance cardiaque. Ce Mini projet a été pour nous l’occasion d’approfondir notre connaissance de langage de programmation C.

Pour conclure, au cours de notre mini projet, nous avons entamé une démarche, à mi-chemin entre la gestion du projet et l’auto-formation, qui a permis de mettre la main sur de nombreuses ressources et composants. Ceci nous a permis de passer toutes les étapes de développement