**Cahier de charge**

**Introduction :**

L'Organisation mondiale de la santé a estimé que 12 millions de décès surviennent dans le monde chaque année en raison de maladies cardiaques. La moitié des décès aux États-Unis et dans d'autres pays développés sont dus à des maladies cardio-vasculaires. Le pronostic précoce des maladies cardiovasculaires peut aider à prendre des décisions sur les changements de mode de vie chez les patients à haut risque et à son tour réduire les complications. Cette recherche vise à identifier les facteurs de risque les plus pertinents des maladies cardiaques et à prédire le risque global à l'aide de la régression logistique et préparation des données.

**Tâche :**

La tâche est de prédire si le patient a un risque de 10 ans de maladie coronarienne CHD ou non. En outre, les participants ont également demandé à créer une visualisation des données sur les données pour obtenir des informations exploitables sur le sujet.

**Source :**

La base de données est accessible au public sur le site Web de Kaggle et provient d'une étude cardiovasculaire en cours sur des résidents de la ville de Framingham, Massachusetts. L'objectif de la classification est de prédire si le patient présente un risque de maladie coronarienne (CHD) à 10 ans. L'ensemble de données fournit des informations sur les patients. Il comprend plus de 4 000 observations et 15 variables.

**Variables :**

Chaque variable est un facteur de risque potentiel. Il existe à la fois des facteurs de risque démographiques, comportementaux et médicaux. Description des données Démographique :

• Sexe : homme ou femme ("M" ou "F")

• Age : Âge du patient (Continu - Bien que les âges enregistrés aient été tronqués en nombres entiers, le concept d'âge est continu) Comportementale

• Is\_smoking : si le patient est ou non un fumeur actuel ("YES" ou "NO")

• CigsPerDay : le nombre de cigarettes que la personne a fumé en moyenne en une journée (peut être considéré comme continu car on peut avoir n'importe quel nombre de cigarettes, même une demi-cigarette). Antécédents médicaux)

• BPMeds : que le patient prenne ou non des médicaments contre l'hypertension (valeur nominale)

• preleventStroke : si le patient a déjà eu un accident vasculaire cérébral (nominal)

• prevalentHyp : si le patient était ou non hypertendu (nominal)

• Diabetes : si le patient était ou non diabétique (nominal)

**Médical (actuel)**

• totChol: taux de cholestérol total (en continu)

• sysBP: tension artérielle systolique (continue)

• diaBP: tension artérielle diastolique (continue)

• BMI: indice de masse corporelle (en continu)

• hearteRate: fréquence cardiaque (continue - Dans la recherche médicale, des variables telles que la fréquence cardiaque sont en fait discrètes, mais sont considérées comme continues en raison du grand nombre de valeurs possibles.)

• Glucose: niveau de glucose (en continu)

**Variable de prédiction (cible souhaitée)**

• TenYearCHD : 10 ans de risque de maladie coronarienne CHD

(binaire : «1», signifie «Yes», «0» signifie «No»).

1. **Dataset à utiliser :**

https://www.kaggle.com/christofel04/cardiovascular-study-dataset-predict-heart-disea

1. Implémentation en java.