

Analyse des données cliniques relatives au diagnostic du cancer du sein

CONTEXTE

Le jeu de données mis à votre disposition contient des données cliniques relatives au diagnostic du cancer du sein. Les caractéristiques (variables) sont calculées à partir d'une image numérisée d'une aspiration à l'aiguille fine (FNA) d'une masse mammaire. Elles décrivent les caractéristiques des noyaux cellulaires présents dans l'image.

Description du jeu de données

Le jeu de données mis à votre disposition pour faire cette analyse est constitué de d'un fichier csv contenant les caractéristiques décrites plus haut. L'ensemble des variables est décrit sur le site dont le lien est le suivant :

<https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29>

TRAVAIL A FAIRE

Il vous est demandé de développer un réseau de neurones artificiels (ANN) pour prédire si une tumeur donnée est bénigne (B) ou maligne (M) en fonction des caractéristiques données.

- La variable cible est la variable ' **diagnosis** '
- Utiliser toutes les variables indépendantes pour entraîner le réseau de neurones sauf les variables **id** et **Unnamed :32**
- Prendre le soin de normaliser les variables indépendantes
- Transformer la variable cible ' **diagnosis** ' en variable numérique (M=1 et B=0)

Les résultats des travaux seront transmis par email à l'adresse ibrahimkone96@gmail.com au plus tard le vendredi 07 juin 2024.