**PREDICTION CANCER DU SEIN**

**I – Identifier et définir le problème**

Le problème est : Pouvons-nous détecter des cellules cancéreuses sur une coupe d’échantillon de tumeur afin de localiser précisément les zones cancéreuses et ainsi évaluer la gravité du cancer ?

Pour répondre à cette question, le modèle va prendre comme entrée une image provenant de la base de données, et en sortie il dira s’il y a cancer ou non. Il s’agit donc d’une classification.

**II – Comprendre les données**

* Charger la base de données dans un tableau

La base de données que nous avons choisie est constituée de 277 524 photos dont 198 738 où il n’y a pas de cellules cancéreuses et 78 786 où il y en a. Cette base de donnée a été créé avec les photos de 279 patients, pour chaque patient il y a un dossier avec deux sous dossiers : le dossier 0 qui contient les photos non cancéreuses et le dossier 1 où il y en a.

Pour pouvoir utiliser les données, nous les chargeons dans un tableau. Ce tableau est composé d’une colonne avec l’image (c’est à dire un vecteur à 3 dimensions car les images sont en couleur) et d’une deuxième colonne avec le diagnostic, c’est à dire 1 si la photo est cancéreuse et 0 sinon.

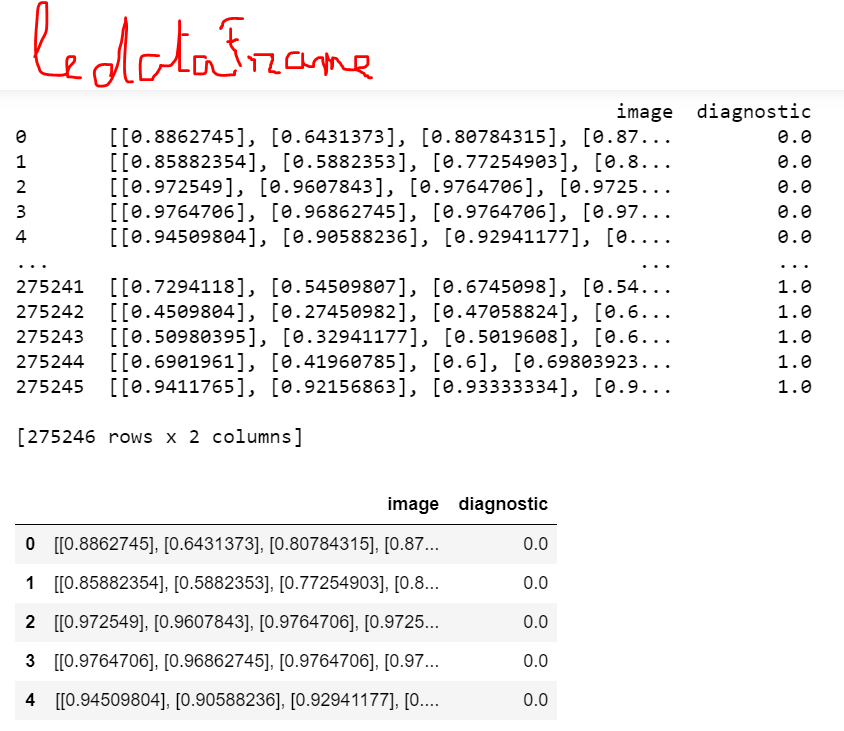
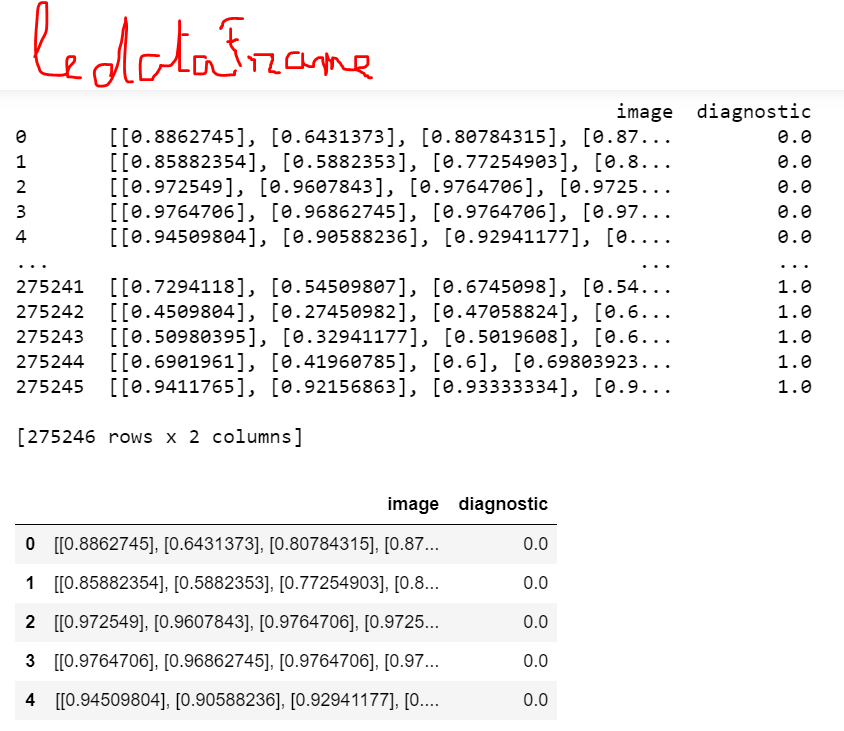
* Sélectionner les photos

Lors de la création de notre fichier, nous nous sommes rendues compte que toutes les photos n’avaient pas la même taille. En effet, la plupart des photos sont de la taille 50x50 mais certaines mesurent 50x48.

Pour résoudre ce problème, nous avons choisi de réduire toutes les images à la même taille : 50x48 (en pratique, on enlève la dernière colonne à toutes les images de taille 50x50). Après avoir fait ça, nous nous sommes rendu compte qu’il restait une infime partie des images qui étaient encore plus petites, nous avons donc décidé de les ignorées et de ne pas les gardées dans notre tableau.

* Transformer la base de données en dataframe

Une fois le tableau créer, nous avons créé un dataframe, ce qui va nous permettre d’accéder aux données rapidement et ?????????????????



* Créer un fichier de sauvegarde .npy

**III – Choisir une méthode**

**IV – choisir un modèle et comprendre les grandes lignes**

(Être capable en 2-3 phrases de faire comprendre à vos collègues) le fonctionnement du modèle et le mécanisme d'apprentissage sous-jacent

**V – Appliquer le modèle et commenter les résultats**