***Projet Machine Learning***

***Presentation :***

**Diabète :**

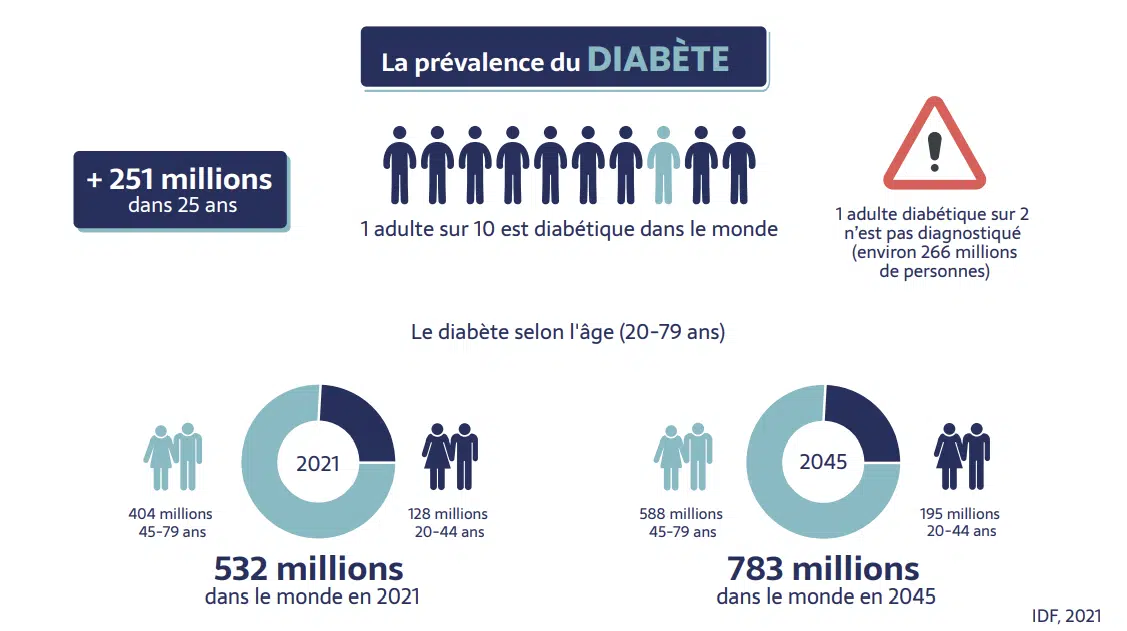
Le diabète est une maladie chronique qui apparaît lorsque le pancréas ne produit pas suffisamment d’insuline ou que l’organisme n’utilise pas correctement l’insuline qu’il produit. L’insuline est une hormone qui régule la concentration de sucre dans le sang.

L’hyperglycémie, ou concentration sanguine élevée de sucre, est un effet fréquent du diabète non contrôlé qui conduit avec le temps à des atteintes graves de nombreux systèmes organiques et plus particulièrement des nerfs et des vaisseaux sanguins.

Les aliments sont composés de lipides (graisses), protéines (protéines animales ou végétales) et glucides (sucres, féculents). Ce sont eux qui fournissent l’essentiel de l’énergie dont a besoin le corps pour fonctionner, passent dans l’intestin, puis rejoignent la circulation sanguine. Le pancréas détecte l’augmentation de la glycémie. Les cellules bêta du pancréas, regroupées en amas appelés îlots de Langerhans, sécrètent de l’insuline.

Insuline fonctionne comme une clé, elle permet au glucose de pénétrer dans les cellules de l’organisme où il va pouvoir être transformé et stocké le glucose diminue alors dans le sang.

En 2014, 8,5% de la population adulte (18 ans et plus) était diabétique. En 2015, le diabète a été la cause directe de 1,6 million de décès et en 2012 l'hyperglycémie avait causé 2,2 millions de décès supplémentaires.



***Le choix des methodes :***

Dans ce cas nous avons travaillé avec :

Arbre de décision : algorithme d'apprentissage supervisé

K-means : est un  [algorithme non supervisé](https://mrmint.fr/lapprentissage-non-supervise-machine-learning)  de **clustering**

***Présentation des methodes :***

***Arbre de décision :***

Les arbres de décision sont une méthode d'apprentissage automatique pour la classification ou la régression. Cela fonctionne en segmentant l'ensemble de données via des instructions de contrôle if-else appliquées aux fonctionnalités.

Un arbre de décision est une structure arborescente semblable à un organigramme ou un nœud interne représente une caractéristique (ou un attribut), la branche représente une règle de décision et chaque nœud feuille représente le résultat, cette structure aide pour prendre la décision.

Les arbres de décision peuvent être facilement visualisés dans un graphique arborescent qui facilite encore plus la compréhension et l'interprétation du modèle.

***K-means :***

Le cluster est défini comme des groupes de points de données tels que les points de données d'un groupe seront similaires ou liés les uns aux autres et différents des points de données d'un autre groupe. Et le processus est connu sous le nom de regroupement. L'objectif du clustering est de déterminer le groupement intrinsèque dans un ensemble de données non étiquetées. Le clustering est une forme d'apprentissage non supervisé car il ne nécessite pas de données étiquetées.

K-means est un algorithme de clustering partitionnel non supervisé qui est basé sur le regroupement de données en k - nombre de clusters en déterminant le centroïde à l'aide de la méthode euclidienne ou Manhattan pour le calcul de la distance. Il regroupe l'objet en fonction de la distance minimale.

🡪 Notre dataset est composée de 2000 lignes et 9 colonnes

La colonne outcome est la caractéristique que nous allons prédire, 0 signifie « pas de diabète » 1 signifie « diabétique »

Sur cette base il y a 684 étiqueté comme 1 et 1316 comme 0