**Sujet Projet FDD**

**Membres du groupe :**

- Olivier Chikuru Buhendwa ([olivierchikuru\_buhendwa@ens.univ-artois.fr](mailto:olivierchikuru_buhendwa@ens.univ-artois.fr))

- Duc Viet Nguyen ([ducviet\_nguyen@ens.univ-artois.fr](mailto:ducviet_nguyen@ens.univ-artois.fr))

- Théo Gayant ([theo\_gayant@ens.univ-artois.fr](mailto:theo_gayant@ens.univ-artois.fr))

**Présentation du sujet :** Prédiction du diabète chez les femmes

Le [diabète](https://www.who.int/fr/news-room/fact-sheets/detail/diabetes) est une maladie chronique qui survient lorsque le pancréas ne produit pas suffisamment d’insuline ou lorsque l’organisme n’est pas capable d’utiliser efficacement l’insuline qu’il produit. L’insuline est une hormone régulatrice de la glycémie. L’hyperglycémie (la concentration trop élevée de glucose dans le sang) est un effet fréquent du diabète non équilibré, qui entraîne avec le temps des atteintes graves de nombreuses parties de l’organisme et plus particulièrement des nerfs et des vaisseaux sanguins.

**Le jeu de donnée :**

Cet ensemble de données provient à l'origine d’un échantillon publié par un Institut du diabète.

L'objectif de l'ensemble de données est de prédire avec un diagnostic si un patient est atteint du diabète. Plusieurs contraintes ont été placées sur la sélection de ces instances dans une base de données plus grande. En particulier, tous les patients ici sont des femmes âgées d'au moins 21 ans.

Les ensembles de données comprennent plusieurs variables prédictives médicales et une variable cible, « Outcome ». Les variables prédictives comprennent le nombre de grossesses que la patiente a eues, son IMC, son taux d'insuline, son âge, etc.

Chaque ligne représente une patiente et les colonnes sont :

* Grossesses: nombre de fois que la patiente était enceinte
* Glucose: concentration en glucose plasmatique 2 heures dans un test de tolérance au glucose par voie orale
* BloodPressure: pression artérielle diastolique (mm Hg)
* SkinThickness: épaisseur du pli cutané des triceps (mm)
* Insuline: insuline sérique de 2 heures (mu U / ml)
* IMC: indice de masse corporelle (poids en kg / (taille en m) ^ 2)
* DiabetesPedigreeFunction: Fonction pedigree du diabète
* Age: Age (ans)
* Résultat: Variable de classe (0 ou 1).

**Questions :**

* Quel âge moyen, une femme a plus de chance d’être diabétique ?
* Le nombre de grossesses a-t-il une incidence sur le pourcentage d’être diabétique ?
* Peut-on prédire le diabète d’une patiente avec les attributs que nous avons à disposition ?
* L’IMC a t-elle un impact sur le fait d’être diabétique ?
* La tension artérielle des patientes diabétiques est relativement élevée à un certain nombre ?