

---

## Série – Naive Bayes

---

### Exercice 1

L'objectif est d'entraîner un modèle Naive Bayes pour prédire la maladie de diabète chez des patients. Le dataset est : '*pima-indians-diabetes.data.csv*'. Pour ceci exécutez les tâches suivantes et répondez aux questions ci-dessous :

1. Chargez le fichier de données '*pima-indians-diabetes.data.csv*' dans un dataframe data.
2. Divisez le dataframe data en deux parties X et Y: X pour les données input (variables descriptives) et Y pour les données Target (la classe)
3. Divisez vos données X et Y en 80% de données d'entraînement (`x_train`, `y_train`) et en 20% de données de test (`x_test`, `y_test`)
4. Créez un modèle Naive Bayes (`GaussianNB`) et ajustez-le aux données d'entraînement (garder les paramètres par défaut).
5. Testez votre modèle sur les données de test. Quelles sont les prédictions pour les 5 premiers patients ?
6. Quelle est l'accuracy de votre modèle ?
7. Dressez la matrice de confusion et le rapport de classification de votre modèle.
8. Testez votre modèle avec la méthode de validation croisée. Quelles sont les nouvelles valeurs d'accuracy ? Quelle est la valeur moyenne ?

## Exercice 2

Le tableau suivant contient des renseignements sur des prêts et peut être utilisé pour tenter de prédire si un emprunteur sera en défaut (la dernière colonne est la classe).

Tid	Home Owner	Marital Status	Annual Income	Default Borrower
1	Yes	Single	High	No
2	No	Married	High	No
3	No	Single	Low	No
4	Yes	Married	High	No
5	No	Divorced	Low	Yes
6	No	Married	Low	No
7	Yes	Divorced	High	No
8	No	Single	Low	Yes
9	No	Married	Low	No
10	No	Single	Low	Yes
11	Yes	Married	Low	Yes

On veut déterminer si  $X = (\text{Home Owner} = \text{No}, \text{Marital Status} = \text{Married}, \text{Income} = \text{High})$  doit être classé comme emprunteur en défaut (Defaulted Borrower) ou non.

1. Rappelez brièvement le principe de l'algorithme Naive Bayes.
2. Appliquez l'algorithme Naive Bayes pour déterminer la classe de X.