

Projet : Étude comparative des modèles de classification

Contexte

Cet ensemble de données provient de l'Institut national du diabète et des maladies digestives et rénales. L'objectif de l'ensemble de données est de prédire diagnostiquement si un patient souffre ou non de diabète, sur la base de certaines mesures diagnostiques incluses dans l'ensemble de données. Plusieurs contraintes ont été imposées sur la sélection de ces instances à partir d'une base de données plus grande. En particulier, tous les patients ici sont des femmes âgées d'au moins 21 ans et d'origine indienne Pima.

Contenu

*L'ensemble de données se compose de plusieurs variables prédictives médicales et d'une variable cible, **Outcome**. Les variables prédictives incluent le nombre de grossesses que la patiente a eues (**Pregnancies**), son IMC (**BMI**), son taux d'insuline (**Insulin**), son âge (**Age**), etc.*

Travail à faire : Prédire avec précision si les patients de l'ensemble de données sont diabétiques ou non ?

Après l'importation du jeu de données "diabetes.csv", sous Python, mettez en place les modèles de classification ci-après :

- *Regression Logistique*
- *Arbre de classification*
- *Naïve Bayes*
- *SVM (Support Vector Machine)*

Faire une étude comparative de ces différents modèles (vous utiliserez les métriques telles que : la Précision, Recall, F1-score et Support) et tirez-en une conclusion.