

Projet : Partie 4

Classification et Prédiction

La classification est une forme d'analyse de données qui extrait des modèles décrivant des classes de données importantes. Ils permettent d'étiqueter de nouvelles données en se basant sur des exemples déjà vus. La classification a d'ailleurs de nombreuses applications, notamment la détection des fraudes (classes: YES/NO), le marketing ciblé (Classes: Standard/Premium/Business), la prédiction des performances (Classes: Good/Medium/Bad), le diagnostic médical (Classes: Concer/Corona/Pneumonia/...), etc.

De nombreuses méthodes ont été proposées afin de développer des techniques de classification et de prédiction capables de gérer de grandes quantités de données. Chacune de ces méthodes comporte des avantages et des inconvénients la rendant plus ou moins adaptée à un type de dataset. Ainsi, dans cette quatrième partie du projet, il vous est demandé de construire un classificateur permettant de prédire si une instance représentant un employé d'une entreprise risque de quitter son job (Yes/No). Soit l'attribut "*Attrition*" du **dataset 1**. Il vous est alors demandé de :

- A. Application des algorithmes de classification:
 - a. Séparer le dataset en données d'apprentissages et données de tests (80% par classe / 20% par classe, respectivement).
 - b. Programmer les deux algorithmes de classification "**Decision Trees**" et "**Random Forest**".
 - c. Appliquer les deux algorithmes sur les instances du dataset.
 - d. Illustrer par des exemples.
 - e. Donner la Matrice de confusion.
 - f. Évaluer et Comparer les modèles de classification en calculant les mesures : ACCURACY, SENSITIVITY, SPECIFICITY, PRÉCISION, RAPPEL, F-SCORE pour chaque classe & globale en plus du temps moyen d'exécution.
- B. Options avancées de l'IHM:
 - a. Choix de la méthode de DataMining à exécuter.
 - b. Insertion des données sur un employé et en déduire sa classe.

Notes:

- Chaque binôme est tenu d'envoyer la version électronique du rapport au plus tard: **Samedi 10 Décembre 2022 à 23h59**.
- Chaque binôme devra présenter son interface et code source le **Dimanche 11 Décembre 2022** durant la séance de TP.

Bon courage !