Problème II : Arbres de décision, une application sur des données réelles :

Dans cet exercice, il s'agit d'expérimenter une implémentation de l'arbre de décision pour un problème de classification binaire.

1. A partir du répertoire en ligne "Échantillons de données" de l'ENT, choisissez un échantillon de données (Le choix de l'échantillon de données doit être communiqué à l'enseignant et validé avant de commencer le travail). Commencez par effectuer une analyse exploratoire descriptive de votre base de données (typologie des variables, valeurs manquantes, distributions des observations pour les différentes variables, boxplots, etc.).
2. Utilisez une implémentation de l'arbre de décision sous R (ou autre) pour construire un classifieur ayant pour objectif de prédire la classe de la variable dépendante dans votre base de données après avoir observé les variables indépendantes, en respectant les consignes suivantes :
   (a) Optimisez votre arbre de classification en passant par des techniques vues dans le cours (par exemple, l'optimisation des hyperparamètres par validation croisée, élagage, etc.).
   (b) Selon le mode opératoire, visualisez l'arbre optimal généré.
3. Vous serez évalué sur les résultats de votre modèle optimal **appliqué aux données test** :
   (a) Explorer le résultat de la classification :
      i. Detailed Accuracy By Class (Precision, Recall,...)
      ii. Confusion Matrix

iii. Etc. (Je vous invite à calculer autant de métriques de performance que possible tout en justifiant les résultats obtenus à chaque fois.)

(b) Veuillez inclure dans le compte rendu la courbe ROC et l'AUC de votre modèle optimal, ainsi que le code sous R correspondant.

(c) Donnez une conclusion/interprétation globale par rapport aux résultats obtenus. N'hésitez pas à faire preuve de créativité et à aller au-delà des questions demandées. Toute idée d'analyse à valeur ajoutée sera fortement appréciée. ?

Problème IV : Foret aléatoire, une application sur des données réelles :

Dans cet exercice, il s'agit d'expérimenter une implémentation d'une forets aléatoires de décision.

1. A partir du répertoire en ligne "Échantillons de données" de l'ENT, considérez le même échantillon de données du problème II. Utiliser différentes implémentations (au moins deux) du foret d'arbres de décision sous R et pour chaque implémentation Présentez-moi **avec clarté** les étapes de votre raisonnement menant au modèle optimal (Out-of sample estimation, Cross Validation, Élagage, optimisation des hyperparamètres, etc.)
2. Analyser et comparer les erreurs de la classification (Detailed Accuracy By Class (Precision, Recall,...), Confusion Matrix, Courbes ROC, AUC, etc.)
3. Interpréter les performances des différentes implémentations des forets aléatoires des parties (1 et 2) .