

Consignes du Projet

Vous réaliserez deux modèles : une régression logistique et un arbre de décision.

Objectif principal : augmenter la valeur des métriques de qualité de votre modèle.

Etapes obligatoires :

Premier cycle :

1. Définition et nomenclature des données
2. Traiter les données aberrantes
3. Présenter les liaisons/corrélations avec la variable cible
4. Entraîner les modèles
5. Présenter et interpréter les résultats des métriques de qualité

Deuxième cycle :

1. Modification des variables **et/ou** de la spécification des modèles
2. Entraîner les modèles
3. Présenter et interpréter les résultats des métriques de qualité
4. Comparer la variation avec les performances précédentes

Troisième cycle :

1. Présenter et justifier la version finale de vos modèles
2. Utiliser les données **nouvelle_data.csv** pour prédire les classes de revenus

Conditions et méthodes obligatoires :

1. Vous répétez le **deuxième cycle au moins deux fois** pour augmenter la performance de votre modèle (*i.e* une augmentation de la valeur des métriques de qualité)
2. Vos deux modèles, entre eux, **doivent toujours être entraînés sur les mêmes données et selon les mêmes variables** afin de pouvoir comparer leurs performances
3. Vos métriques de qualité doivent être composées **à minima de l'AUC, du gAUC, du F1-Score et de l'Accuracy**

Livrables obligatoires :

1. Un rapport technique de vos méthodes au format .pdf
2. Un document .html restituant le code utilisé

Mises en forme obligatoires :

Aucune partie du travail ne doit être vulgarisée. Faites correspondre les étapes de votre code aux étapes expliquées dans votre rapport technique. Soyez méthodiques et structurez vos documents pour garantir la clarté et la précision de vos explications.

Conseils :

Numérotez les versions de vos données et les versions de vos modèles et indiquez clairement à quels moments ces versions interviennent.