

TP1 : Prétraitement des données en apprentissage automatique

Objectif : Préparer le dataset « Titanic » pour une analyse en apprentissage automatique en appliquant les différentes étapes du nettoyage et de la transformation des données.

Étape 1 : Chargement et Exploration du Dataset

Questions :

1. Charger le dataset dans un DataFrame Pandas.
2. Afficher les 5 premières lignes du dataset.
3. Vérifier les informations générales (types de données, valeurs nulles, etc.).
4. Afficher les statistiques descriptives du dataset.

Étape 2 : Nettoyage des Données

Questions :

5. Combien de valeurs manquantes y a-t-il dans chaque colonne ?
6. Supprimer les lignes où des valeurs importantes sont manquantes (ex. : "embarked").
7. Remplacer les valeurs manquantes de "age" par la moyenne.
8. Vérifier et supprimer les doublons s'il y en a.

Étape 3 : Analyse Exploratoire et Visualisation

Questions :

9. Afficher un histogramme de la répartition des âges.
10. Afficher un diagramme en barres du nombre de survivants par classe.

Étape 4 : Encodage des Variables Catégorielles

Questions :

11. Convertir la colonne "sex" en numérique (0 = male, 1 = female).
12. Convertir la colonne "embarked" en variables binaires (One-Hot Encoding).

Étape 5 : Mise à l'Échelle des Données**Questions :**

13. Appliquer la normalisation (Min-Max Scaling) sur age et fare.
14. Appliquer la standardisation (Z-score) sur age et fare.

Étape 6 : Séparation en Ensembles d'Entraînement et de Test**Questions :**

15. Séparer les données en X (features) et y (target : "survived").
16. Diviser les données en 70% entraînement et 30% test.