

projet-formation

Atika El hadji

02 December 2024

Partie 1: Data visualisation avec le dataset Titanic

1 Préparation des données (Pandas)

- Importation de données
- Présentation du dataset Titanic
- Nettoyage des données : gestion des valeurs manquantes et et données aberrantes
- analyse initiale (statistiques descriptives)

2 Visualization des données (Matplotlib, seaborn)

- Création de graphiques de base :histogrammes , digramme en barre.
- visualisation des relations entre les variables (matrice de corrélation , diagrammes de dispersion).

Partie 2: Les algorithmes de machine learning

3 Préparation des données pour ML

- du machine learning et types d'apprentissage (supervisé, non supervisé).
- Présentation des algorithmes supervisés : classification et régression.
- Splitting des données : jeu d'entraînement et de test.

4 Modèles de Classification sur le Dataset Titanic

- Prétraitement des données : encodage des variables catégorielles, normalisation.
- Construction des modèles.
- Évaluation des performances de chaque modèle
- Faire une comparaison entre les modèles utilisées et choisir le meilleur