

Université Mohammed Premier
École Nationale des Sciences Appliquées
Filière Génie Informatique

TP : Machine Learning

Partie 4 : Régression logistique

L'objectif de l'exercice est de prédire l'acceptation ou non des crédits particuliers, sur la base de certaines mesures incluses dans le Dataset.

1. L'ensemble de données 'dataset.csv', se composent de plusieurs variables prédictives et d'une variable cible "Loan_Status". Charger le dataset dans un DataFrame Pandas.
2. Nettoyez le jeu de données avec Pandas/sklearn. En supprimant ou en remplaçant les lignes avec des valeurs manquantes.
3. Normaliser les valeurs numériques
4. Encoder les valeurs catégorielles
5. Séparer la colonne cible du reste des colonnes
6. Divisez le dataset en deux 80% pour le train_set et le reste pour le test_set.
7. Construisez un modèle de régression logistique qui permet la classification des demandes de crédit.
8. Evaluer votre modèle à l'aide de la matrice de confusion. Afficher puis interpréter la matrice.
9. Mesurez la précision, recall, et le score F1 du votre modèle. Ecrivez un paragraphe qui permet d'interpréter vos résultats.
10. Afficher la courbe ROC et évaluer le score AUC obtenu.
11. Récupérer le seuil optimal de régression à l'aide de la courbe ROC.