



TD - Azure Machine Learning

Exercice 1 - Apprentissage supervisé

1. Lancez le logiciel **AZURE Machine Learning Studio** et créez un nouveau projet que vous intitulez "TD Azure".
2. Créez une nouvelle expérience que vous nommerez "Exercice 1 - Apprentissage supervisé" et affectez-la au projet TD Azure.
3. Importez le fichier de données "Adult Census Income.txt" (+New -> Dataset). Sélectionnez le type ".tsv" lors de l'importation.
4. Glissez/déposez le jeu de données dans votre expérience. Visualisez les données. Combien y-a-t'il d'observations et de variables ? De quel type ?
5. Affichez les statistiques descriptives de l'ensemble des variables (**Summarize Data**).
6. Quelles sont les variables susceptibles d'être retirées pour l'analyse ? Pourquoi ? Otez-les de l'analyse.
7. Partitionnez aléatoirement les données en échantillons d'apprentissage (70%) et de test (30%).
8. Établissez une régression logistique et appliquez-la sur l'échantillon test en sélectionnant la variable cible (**Train Model**).
9. Calculez et visualisez les scores.
10. Évaluez le modèle et commentez la matrice de confusion.
11. Comparez ce modèle avec celui des arbres de décision. Sur le même principe que la question 8, établissez un arbre de décision et appliquez-le sur l'échantillon test.
12. Quel est selon vous le meilleur modèle ? Pourquoi ?

Exercice 2 - Apprentissage non supervisé

1. Créez une nouvelle expérience que vous nommerez "Exercice 2 - Apprentissage non supervisé" et affectez là au projet TD Azure.
2. A partir du même jeu de données, recodez la variable cible en binaire (0 et 1). Cette étape est à réaliser en 2 modules.
3. Normalisez les données en sélectionnant les données quantitatives.
4. Réalisez un K-means et appliquez-le aux données (**Train Clustering Model**).
5. Afin d'obtenir la variable d'affectation aux clusters, utilisez le module **Select Columns in Dataset**.

Exercice 3 - Apprentissage non supervisé ACP (Analyse en Composante Principale)

1. Créez une nouvelle expérience que vous nommerez « Exercice 3 – ACP » et affectez-la au projet TD Azure.
2. Importez le jeu de données « Villes universitaires_MV.txt ».
3. Nettoyer les données manquantes du fichier de données.
4. Normaliser les données avant de réaliser une ACP. Peut-on tirer une conclusion sur les différentes villes universitaires ?