



## EXPLOITATION DES DONNÉES ET MODÈLES D'APPRENTISSAGE EN DATA SCIENCE

### COMPÉTENCE PRINCIPALE VISÉE

Mettre en oeuvre la chaîne complète d'un projet Data

### OBJECTIFS PÉDAGOGIQUES

L'objectif de la formation est de permettre aux participants de :

- Interagir avec les spécialistes dans un cadre professionnel
- Mettre en œuvre la chaîne complète d'un projet Data
- Identifier les grandes familles de méthodes d'utilisation des données
- Concevoir et organiser l'exploitation des données
- Concevoir et appliquer les méthodes de validation des modèles d'apprentissage
- Interagir avec les données (nettoyage, prévisualisation...)
- Justifier les résultats d'un projet de sciences de données dans un cadre opérationnel

### PUBLIC

Tous collaborateurs en charge de mettre en oeuvre un projet de Data Science

### PRÉREQUIS

- Notions en Data Science
- Connaissances en statistique et/ou en programmation

### CONTENU

#### PARTIE 1 - PREPARATION ET VISUALISATION DES DONNEES

- Accéder aux données
- Importer des données
- Comprendre l'importance du processus de nettoyage des données
- Identifier les caractéristiques d'un ensemble de données
- Nettoyer et préparer des ensembles de données
- Prétraiter des données (normalisation, discrétion)
- Les principes de la visualisation de données
- Réduction de la dimension, analyse en composantes principales (ACP)

Modalités pédagogiques : 50 % de Travaux Pratiques sur Plateforme Data Science basée sur des notebooks Scikit-Learn et Python

#### PARTIE 2 - LES MODELES D'APPRENTISSAGE FONDAMENTAUX

- Modéliser un problème de Data Science
- Ajuster et valider un modèle
- La Machine Learning et les capacités d'apprentissage des machines
- Les différentes familles d'algorithmes
- Apprentissage supervisé
- Construire un modèle de régression linéaire
- Les coefficients de régression
- k-NN
- Arbres de décision
- Validation d'un modèle prédictif
- Apprentissage non supervisé
- Partitionnement en k-means
- Regroupement hiérarchique (clustering)
- Règles d'association

### SESSIONS

Villeurbanne : Du 18/09/23 au 22/09/23 à 12h00

L'ouverture de la session est conditionnée par un nombre minimum de participants.

### DURÉE

4,5 jours (32 heures)

### FRAIS INDIVIDUELS

Coût de la formation (repas inclus) : 4006 € H.T.

### ÉQUIPE PÉDAGOGIQUE

Enseignant chercheur du laboratoire LIRIS de l'INSA LYON


### RENSEIGNEMENTS ET INSCRIPTION

Tel : +33(0) 4 72 43 83 93

Fax : +33 (0)4 72 44 34 24

mail : [formation@insavalor.fr](mailto:formation@insavalor.fr)

Préinscription sur [formation.insavalor.fr](http://formation.insavalor.fr)

 Accueil des personnes en situation de handicap nécessitant un besoin spécifique d'accompagnement : nous contacter à l'inscription

- Détections d'anomalies

Modalités pédagogiques : 50 % de Travaux Pratiques sur Plateforme Data Science basée sur des notebooks Scikit-Learn et Python

### **PARTIE 3 – LES MODELES PREDICTIFS AVANCES**

- Régression non linéaire, régression logistique
- Random forests
- Machines à vecteurs de support (SVM)
- Validation d'un modèle prédictif : validation croisée

Modalités pédagogiques : 50 % de Travaux Pratiques sur Plateforme Data Science basée sur des notebooks Scikit-Learn et Python

## **MÉTHODE PÉDAGOGIQUE**

La démarche pédagogique proposée s'appuiera sur l'alternance d'apports théoriques, d'échanges techniques et d'illustrations autour de cas pédagogiques existants. En complément, elle intégrera de nombreuses mises en situation et de travaux pratiques

## **ÉVALUATION ET RÉSULTATS**

### **Évaluation des acquis de la formation**

Evaluation des acquis des apprenants par auto-examen

### **Évaluation de la formation**

Evaluation du ressenti des participants en fin de formation (Niveau 1 KIRKPATRICK)

### **Résultats de l'évaluation**

Le niveau d'appréciation globale de la thématique est évalué à 4.3/5 par les participants