Python : modélisation et représentation des données

Lien : https://innov-systems.com/formation/python-modelisation-et-representation-des-donnees

DURÉE
5 jours (35h)

RÉFÉRENCE
BSI55

CATÉGORIE
Big Data - Langages

OBJECTIFS DE LA FORMATION

À l'issue de cette formation, vous serez capable de :

- Appréhender le principe de la modélisation selon leur problématique
- Arriver à choisir entre la régression et la classification
- Évaluer les performances prédictives d'un algorithme
- Créer des sélections et des classements dans de grands volumes de données pour dégager des tendances
- Utiliser différentes méthodes d'exploration de données

POUR QUI ?

- Responsables Infocentre
- Utilisateurs
- Gestionnaires métiers de bases de données

Innov Systems

Programme détaillé

1 / Introduction au langage Python

- Les caractéristiques du langage Python
- Pourquoi choisir Python pour l’analyse de données ?
- Installation et configuration
- Philosophie de Python (indentation, objet, etc.)
- Les types de données
- Appels de fonctions et méthodes
- Structures de contrôles (boucle, test, exceptions)
- Structures de données et séquences (tuple, liste, primitives, dict)

2 / Modélisation

- Présentation Jupiter Notebook
- Construction d'un modèle : étapes
- Algorithmes supervisés et non supervisés
- Le choix entre la régression et la classification

3 / Evaluation de modèles : procédures

- Les techniques de ré-échantillonnage
- Test de représentativité des données d'apprentissage
- Mesures de performance des modèles prédictifs
- Matrice de confusion, de coût et la courbe ROC et AUC

4 / Algorithmes supervisés

- Le principe de régression linéaire univariée
- La régression multivariée

Innov Systems

- La régression polynomiale
- La régression régularisée
- Le Naive Bayes
- L’arbre de décision
- La régression logistique
- Les K plus proches voisins

5 / Les algorithmes non supervisés

- Le clustering hiérarchique
- Le clustering non hiérarchique
- Les approches mixtes

6 / Analyse en composantes

- Analyse en Composantes Principales
- Analyse du nuage des individus
- Propriétés des composantes principales
- Analyse du nuage des variables
- Interprétation des résultats
- Analyse des correspondances
- Nuages des profils-lignes et des profils-colonnes, la distance du khi-deux
- Analyse des nuages des profils
- Liens entre les analyses
- Interprétation des résultats
- Analyse des Correspondances Multiples
- Analyse des correspondances sur le tableau disjonctif complet
- Propriétés de l’ACM
- Analyse du tableau de Burt
- Analyse Factorielle pour Données Mixtes
- Classification
- Classification ascendante hiérarchique : principe, stratégies d’agrégation, aides à l’interprétation

Inov Systems

- Méthodes de partitionnement : les méthodes des centres mobiles, les nuées dynamiques
- Classification mixte

7 / Analyse de données textuelles

- Collecte et prétraitement des données textuelles
- Extraction d'entités primaires, d'entités nommées et résolution référentielle
- Étiquetage grammatical, analyse syntaxique, analyse sémantique
- Lemmatisation
- Représentation vectorielle des textes
- Pondération TF-IDF
- Word2Vec

Approche pédagogique

- Support Ecrit et Projection
- Exposés Interactifs, Podcasts et Vidéos
- Brainstorming et Jeux de Rôle
- Cas Pratiques et Labs inclus pour leur impact opérationnel
- Test de Validation des Acquis des Connaissances

Prochaines dates programmées

16 au 20 Nov. 2026	Distanciel
04 au 08 Jan. 2027	Présentiel - Casablanca
01 au 05 Mars 2027	Présentiel - Casablanca
26 au 30 Avr. 2027	Présentiel - Casablanca
21 au 25 Juin 2027	Présentiel - Casablanca

Autres dates possibles sur demande. Contactez-nous pour organiser une session intra-entreprise.

Innov Systems

Réservation & Renseignements

Téléphone : +212 522 247 210
Email : contact@innov-systems.com
Web : https//www.innov-systems.com

Innov Systems — Tous droits réservés

Innov Systems