



## Formation Big Data Analytics avec Python modélisation et représentation des données

N° ACTIVITÉ : 11 92 18558 92

TÉLÉPHONE : 01 85 77 07 07

E-MAIL : inscription@hubformation.com

### Objectifs

- | Identifier le principe de la modélisation statistique
- | Choisir entre la régression et la classification en fonction du type de données
- | Évaluer les performances prédictives d'un algorithme
- | Créer des sélections et des classements dans de grands volumes de données pour dégager des tendances

### Public

- | Responsables Infocentre (Datamining, Marketing, Qualité...), utilisateurs et gestionnaires métiers de bases de données.

### Prérequis

- | Connaissances de base en statistiques ou avoir suivi le stages "Statistiques, maîtriser les fondamentaux".
- | Connaissances de base en Python.

### Programme de la formation

#### Introduction à la modélisation

- | Introduction au langage Python.
- | Introduction au logiciel Jupiter Notebook.
- | Les étapes de construction d'un modèle.
- | Les algorithmes supervisés et non supervisés.
- | Le choix entre la régression et la classification.
- | Travaux pratiques Installation de Python 3, d'Anaconda et de Jupiter Notebook.

#### Procédures d'évaluation de modèles

- | Les techniques de ré-échantillonnage en jeu d'apprentissage, de validation et de test.
- | Test de représentativité des données d'apprentissage.
- | Mesures de performance des modèles prédictifs.
- | Matrice de confusion, de coût et la courbe ROC et AUC.
- | Travaux pratiques Mise en place d'échantillonnage de jeux de données. Effectuer des tests d'évaluations sur plusieurs modèles fournis.

#### Les algorithmes supervisés

- | Le principe de régression linéaire univariée.
- | La régression multivariée.
- | La régression polynomiale.
- | La régression régularisée.
- | Le Naive Bayes.
- | La régression logistique.
- | Travaux pratiques Mise en oeuvre des régressions et des classifications sur plusieurs types de données.

#### Les algorithmes non supervisés

- | Le clustering hiérarchique.

Référence	BDA
Durée	4 jours (28h)
Tarif	2 750 €HT

### SESSIONS PROGRAMMÉES

#### A DISTANCE (FRA)

du 1er au 4 octobre 2024

#### PARIS

du 24 au 27 septembre 2024

#### AIX-EN-PROVENCE

du 1er au 4 octobre 2024

#### BORDEAUX

du 1er au 4 octobre 2024

#### LILLE

du 1er au 4 octobre 2024

#### LYON

du 1er au 4 octobre 2024

#### NANTES

du 1er au 4 octobre 2024

#### SOPHIA-ANTIPOLIS

du 1er au 4 octobre 2024

#### STRASBOURG

du 1er au 4 octobre 2024

#### TOULOUSE

du 1er au 4 octobre 2024

[VOIR TOUTES LES DATES](#)

- | Le clustering non hiérarchique.
- | Les approches mixtes.
- | Travaux pratiques Traitements de clustering non supervisés sur plusieurs jeux de données.

### **Analyse en composantes**

- | Analyse en composantes principales.
- | Analyse factorielle des correspondances.
- | Analyse des correspondances multiples.
- | Analyse factorielle pour données mixtes.
- | Classification hiérarchique sur composantes principales.
- | Travaux pratiques Mise en oeuvre de la diminution du nombres des variables et identification des facteurs sous-jacents des dimensions associées à une variabilité importante.

### **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.
- | Travaux pratiques Explorer le contenu d'une base de textes en utilisant l'analyse sémantique latente.

## **Méthode pédagogique**

Chaque participant travaille sur un poste informatique qui lui est dédié. Un support de cours lui est remis soit en début soit en fin de cours. La théorie est complétée par des cas pratiques ou exercices corrigés et discutés avec le formateur. Le formateur projette une présentation pour animer la formation et reste disponible pour répondre à toutes les questions.

## **Méthode d'évaluation**

Tout au long de la formation, les exercices et mises en situation permettent de valider et contrôler les acquis du stagiaire. En fin de formation, le stagiaire complète un QCM d'auto-évaluation.

---

## **Suivre cette formation à distance**

Voici les prérequis techniques pour pouvoir suivre le cours à distance :

- | Un ordinateur avec webcam, micro, haut-parleur et un navigateur (de préférence Chrome ou Firefox). Un casque n'est pas nécessaire suivant l'environnement.
- | Une connexion Internet de type ADSL ou supérieure. Attention, une connexion Internet ne permettant pas, par exemple, de recevoir la télévision par Internet, ne sera pas suffisante, cela engendrera des déconnexions intempestives du stagiaire et dérangera toute la classe.
- | Privilégier une connexion filaire plutôt que le Wifi.
- | Avoir accès au poste depuis lequel vous suivrez le cours à distance au moins 2 jours avant la formation pour effectuer les tests de connexion préalables.
- | Votre numéro de téléphone portable (pour l'envoi du mot de passe d'accès aux supports de cours et pour une messagerie instantanée autre que celle intégrée à la classe virtuelle).
- | Selon la formation, une configuration spécifique de votre machine peut être attendue, merci de nous contacter.
- | Pour les formations incluant le passage d'une certification la dernière journée, un voucher vous est fourni pour passer l'examen en ligne.
- | Pour les formations logiciel (Adobe, Microsoft Office...), il est nécessaire d'avoir le logiciel installé sur votre machine, nous ne fournissons pas de licence ou de version test.
- | Horaires identiques au présentiel.

---

## **Accessibilité**

Les sessions de formation se déroulent sur des sites différents selon les villes ou les dates, merci de nous contacter pour vérifier l'accessibilité aux personnes à mobilité réduite.

Pour tout besoin spécifique (vue, audition...), veuillez nous contacter au 01 85 77 07 07.