

# Parcours certifiant construire et développer des modèles de big data

## Bloc de compétences d'un titre RNCP

Titre RNCP de 32 jours - 224h

Réf : ZDM - Prix 2024 : 14 900€ HT

Ce parcours de formation représente le quatrième bloc de compétences du titre RNCP de niveau 7 (Bac +5) "Expert en informatique et système d'information - big data" reconnu par l'État. L'ensemble de ces formations vous permettra de comprendre les statistiques descriptives et le traitement des données associées. Vous apprendrez aussi comment gérer un projet big data, le rôle stratégique de la valeur des données pour l'entreprise, la modélisation statistique, les méthodes et solutions de machine learning et de deep learning.

### Ce cycle est composé de :

- Modéliser les processus métiers dans une démarche d'urbanisation (Réf. MPU, 3 jours)
- Statistiques descriptives, introduction (Réf. UES, 2 jours)
- Big Data, méthodes et solutions pratiques pour l'analyse des données (Réf. BID, 5 jours)
- Démarche de gestion de la qualité des données (Réf. MQD, 2 jours)
- Modélisation statistique, l'essentiel (Réf. STA, 2 jours)
- Data Analytics avec Python (Réf. BDA, 4 jours)
- Machine learning, méthodes et solutions (Réf. MLB, 4 jours)
- Data Mining par la pratique (Réf. DMP, 3 jours)
- Text Mining par la pratique (Réf. MMD, 3 jours)
- Deep Learning par la pratique (Réf. DPL, 3 jours)
- Certification construire et développer des modèles de big data (Réf. ZDB, 1 jour)

### OBJECTIFS PÉDAGOGIQUES

À l'issue de la formation l'apprenant sera en mesure de :

Comprendre les enjeux de l'alignement stratégique et de l'urbanisation des SI

Comprendre l'intérêt de la statistique descriptive

Comprendre l'écosystème technologique nécessaire pour réaliser un projet big data

Comprendre le principe de la modélisation statistique

Comprendre les différents modèles d'apprentissage

Connaître les principales méthodes de data mining

Comprendre les méthodes de la statistique textuelle

Mettre en place le deep learning

## LE PROGRAMME

dernière mise à jour : 04/2024

### PARTICIPANTS

Toute personne souhaitant construire et développer des modèles de big data.

### PRÉREQUIS

Être titulaire d'un diplôme de niveau 6 (Bac +3) ou d'un niveau 5 (BAC+2) et 3 ans d'expérience, sous réserve de la validation du dossier de Validation des acquis professionnels (VAP).  
Connaissances de base en Python et en statistiques.

### COMPÉTENCES DU FORMATEUR

Les experts qui animent la formation sont des spécialistes des matières abordées. Ils ont été validés par nos équipes pédagogiques tant sur le plan des connaissances métiers que sur celui de la pédagogie, et ce pour chaque cours qu'ils enseignent. Ils ont au minimum cinq à dix années d'expérience dans leur domaine et occupent ou ont occupé des postes à responsabilité en entreprise.

### MODALITÉS D'ÉVALUATION

Le formateur évalue la progression pédagogique du participant tout au long de la formation au moyen de QCM, mises en situation, travaux pratiques...  
Le participant complète également un test de positionnement en amont et en aval pour valider les compétences acquises.

### MOYENS PÉDAGOGIQUES ET TECHNIQUES

- Les moyens pédagogiques et les méthodes d'enseignement utilisés sont principalement : aides audiovisuelles, documentation et support de cours, exercices pratiques d'application et corrigés des exercices pour les stages pratiques, études de cas ou présentation de cas réels pour les séminaires de formation.
- À l'issue de chaque stage ou séminaire, ORSYS fournit aux participants un questionnaire d'évaluation du cours qui est ensuite analysé par nos équipes pédagogiques.
- Une feuille d'émargement par demi-journée de présence est fournie en fin de formation ainsi qu'une attestation de fin de formation si le stagiaire a bien assisté à la totalité de la session.

### MODALITÉS ET DÉLAIS D'ACCÈS

L'inscription doit être finalisée 24 heures avant le début de la formation.

### ACCESSIBILITÉ AUX PERSONNES HANDICAPÉES

Vous avez un besoin spécifique d'accessibilité ? Contactez Mme FOSSE, référente handicap, à l'adresse suivante psh-accueil@orsys.fr pour étudier au mieux votre demande et sa faisabilité.

### 1) Modéliser les processus métiers dans une démarche d'urbanisation

- Principes, concepts, démarches et enjeux.
- La modélisation des processus.
- Lien avec la démarche d'urbanisation.
- Outils de modélisation des processus.

### 2) Statistiques descriptives, introduction

- Définition.
- Formalisation mathématique.
- Traitement statistique des données à une dimension.
- Variables aléatoires.
- Statistique descriptive à deux dimensions : les tableaux de contingence.
- Cas pratique : utilisation des données des participants.

### 3) Big data, méthodes et solutions pratiques pour l'analyse des données

- Comprendre les concepts et les enjeux du big data.
- Les technologies du big data.
- Gérer les données structurées et non structurées.
- Technique et méthodes big data analytics.
- Datavisualisation et cas d'usage concrets.
- Conclusion.

### 4) Démarche de gestion de la qualité des données

- Introduction.
- Définition de la qualité et cadre méthodologique.
- Démarche gestion de la qualité des données.
- Contrôle qualité et bonnes pratiques.
- Suivi opérationnel de la qualité des données.
- Hausser le niveau de la qualité.

### 5) Modélisation statistique, l'essentiel

- Rappels des fondamentaux de la statistique descriptive.
- Démarche et modélisation d'une analyse statistique.
- Paramètre de position et de dispersion.
- Tests et intervalle de confiance.
- Panorama des outils.

### 6) Data Analytics avec Python

- Introduction à la modélisation.
- Procédures d'évaluation de modèles.
- Les algorithmes supervisés.
- Les algorithmes non supervisés.
- Analyse en composantes.
- Analyse de données textuelles.

### 7) Machine learning, méthodes et solutions

- Introduction au machine learning.
- Procédures d'évaluation de modèles.
- Les modèles prédictifs, l'approche fréquentiste.
- Les modèles et apprentissages bayésiens.
- Machine learning en production.

### 8) Data mining par la pratique

- Le projet de data mining.
- Les techniques de l'approche data mining.
- Les outils statistiques.

- La visualisation des données.
- Analyse de données qualitatives et textuelles.

### 9) Text mining par la pratique

- Les approches traditionnelles en text mining.
- Feature Engineering pour la représentation de texte.
- La similarité des textes et classification non supervisée.
- La classification supervisée du texte.
- Natural Language Processing et deep learning.

### 10) Deep learning par la pratique

- Introduction.
- Introduction aux réseaux de neurones artificiels.
- Entraînement de réseaux de neurones profonds.
- Réseaux de neurones convolutifs.
- Deep learning avec Keras.
- Réseaux de neurones récurrents.
- Autoencodeurs.

## LES DATES

---

Ce parcours est composé d'un ensemble de modules. Les dates indiquées ci-dessous correspondent aux premières sessions possibles du parcours.

**CLASSE À DISTANCE**  
2024 : 28 oct.

**PARIS**  
2024 : 21 oct.