

# DATA ANALYST

université  
PARIS-SACLAY

FACULTÉ  
DES SCIENCES  
D'ORSAY

Formation Continue & VAE  
Se former tout au long de la vie

Diplôme d'université

PARIS SACLAY

créé en 2019

100 % de diplômés,  
100 % de taux de satisfaction



Année universitaire  
2023- 2024



## Présentation

### OBJECTIFS DE LA FORMATION

- ✓ Développer son esprit d'analyse
- ✓ Disposer de bonnes qualités rédactionnelles
- ✓ Savoir communiquer
- ✓ Programmer efficacement
- ✓ Savoir manipuler une base de données
- ✓ Maîtriser des outils de visualisation
- ✓ Modéliser des problèmes orientés données
- ✓ Apprendre à transformer les données brutes
- ✓ Trier les informations et les exploiter pour sortir des indicateurs concrets

### OBJECTIFS PROFESSIONNELS

#### Devenir Analyste de données

- ✓ Acquérir des compétences particulières en informatique et maîtriser des outils spécifiques au Big Data, aussi bien au niveau du stockage des données que de leur traitement (par exemple MapReduce, Hadoop ou encore Spark).
- ✓ Apprendre à dégager des tendances pouvant aboutir à des recommandations sur les stratégies à adopter
- ✓ Maîtriser les outils et méthodes statistiques.

## Responsable de la formation

**Dominique Quadri- Professeure à l'université PARIS SACLAY**

@ dominique.quadri@universite-paris-saclay.fr

## Publics concernés

Cette formation s'adresse à des salariés ou à des adultes en reprise d'études. De niveau BAC +4 minimum dans un domaine scientifique, et justifiant d'une expérience professionnelle permettant de suivre cette formation.

### Dates

Du 01 Septembre à 30 juin 2023

200 heures de formation

### Lieu

À la faculté de sciences

Ou sur site de l'entreprise

# Programme

## Algorithmique et langages dynamiques

25h sept-22

Outils et techniques algorithmiques de pointe : algorithmique sur les graphes (tri topologique, etc.), les techniques de mémorisation, de programmation dynamique et de backtracking. Notion de flots et les algorithmes de calcul de flot maximal

## Data mining / Data warehouse

25h sept-22

Entrepôt de données (ou base de données décisionnelle, ou encore data warehouse) : désigne une base de données utilisée pour collecter, ordonner, journaliser et stocker des informations provenant de base de données opérationnelles et fournir ainsi un socle à l'aide à la décision en entreprise. Présentation des mécanismes de base de ces entrepôts de données et étude des différentes fonctionnalités

## Ecosystème : Hadoop, MapReduce, Spark

25h oct-22

Présenter et manipuler des environnements de programmation distribuée tels que Hadoop. Aborder le patron d'architecture MapReduce sur lequel repose Hadoop. Etudier Spark qui est un framework de calcul distribué qui lui travaille en mémoire vive et non sur MapReduce.

## Maths / Statistiques / Probabilités

25h oct-22

Espace de probabilité discret : notion de modèle d'une expérience aléatoire, construction d'un espace de probabilité, espace et probabilité produit, indépendance et probabilités conditionnelles. Variables aléatoires discrètes : espérance et variance, lois et représentations graphiques, lois et modèles classiques. Statistiques : Loi des grands nombres et problèmes d'estimation. Introduction aux tests d'hypothèses et intervalles de confiance

## Machine Learning

25h nov-22

Bases indispensables en apprentissage automatique ou "Machine Learning" : principales familles de modèles et algorithmes associés (inférence et apprentissage)

## Modélisation

25h nov-22

Espace de probabilité discret : notion de modèle d'une expérience aléatoire, construction d'un espace de probabilité, espace et probabilité produit, indépendance et probabilités conditionnelles. Variables aléatoires discrètes : espérance et variance, lois et représentations graphiques, lois et modèles classiques. Statistiques : Loi des grands nombres et problèmes d'estimation. Introduction aux tests d'hypothèses et intervalles de confiance.

## Technologies mobiles

25h dec-22

Nouvelles technologies mobiles, connaissances fondamentales et compréhension des avantages et inconvénients des différentes solutions existantes, en s'intéressant tout particulièrement au contexte IoT.

## Visualisation des données

25h dec-22

L'enjeu de la représentation des connaissances est de permettre d'explicitier des connaissances humaines de toutes sortes dans un formalisme interprétable par une machine, c'est-à-dire lui permettant de raisonner sur ces connaissances pour remplir des tâches variées (recherche d'information, surveillance et diagnostic de systèmes complexes, ...). Ce cours présentera différents formalismes de représentation de connaissances.

# Format

**Durée de la formation : 200 heures**

**Rythme de la formation : une semaine (25h en présentiel) sur deux, pendant huit semaines et demi.**

**Format des cours : présentiel**



## Modalités de contrôle des connaissances

CONTRÔLE CONTINU INTEGRAL

## Candidature et inscription

Inscription sur dossier et sur entretien après l'avis favorable de la responsable de la formation

**Date ou période d'inscription : date butoir mi-juillet 2023**

### • MODALITÉS DE CANDIDATURE

- Examen du dossier de candidature et entretien de motivation
- Pièces à joindre au dossier de candidature : CV et lettre de motivation

Envoi du dossier avant le **13 juillet** à : @ dominique.quadri@universite-paris-saclay.fr

**université  
PARIS-SACLAY**

**FACULTÉ  
DES SCIENCES  
D'ORSAY**

**Formation Continue & VAE**  
Se former tout au long de la vie

### FRAIS DE FORMATION

Auprès du Service de Formation Continue & VAE,  
Faculté de Sciences d'Orsay- Université PARIS-SACLAY

Chantal ROULET

[Fc.Sciences@universite-paris-saclay.fr](mailto:Fc.Sciences@universite-paris-saclay.fr)

4120 €\*

### FRAIS UNIVERSITAIRES

Auprès du Service de la Scolarité,  
Faculté de Sciences d'Orsay- Université PARIS-SACLAY

380 €

\* Les tarifs ne sont pas assujettis à la TVA