



EXECUTIVE EDUCATION

EXECUTIVE MASTER
STATISTIQUE ET BIG DATA



Le programme dédiés aux futurs Data Scientists souhaitant traiter les problématiques d'entreprise liées au big data

Introduction

Face à l'émergence du phénomène des big data, de nouvelles perspectives s'ouvrent pour les entreprises qui souhaitent tirer profit de l'énorme quantité de données qu'elles collectent quotidiennement. Mais pour constituer une aide à la décision, la valorisation de ces données brutes et multiformes est indispensable. Les métiers de data scientist, qui répondent à cet objectif, sont donc actuellement très recherchés.

Le métier du Data Scientist requiert les compétences pointues pour savoir modéliser, analyser et traiter les données.

L'objectif de ce diplôme universitaire est d'offrir une formation solide de mathématiques appliquées qui couvre toutes les problématiques de nature statistique que l'on peut rencontrer en entreprise. Le but de ce cursus est également d'illustrer l'utilisation des techniques idoines pour le traitement des

problématiques modernes en marketing, en actuariat et en finance.

Au-delà de l'acquisition classique du bagage scientifique du data scientist, cette formation met l'accent à la fois sur la rigueur indispensable pour le traitement des données et sur la présentation des méthodologies statistiques récemment développées. Ces dernières ont pour fonction de produire les outils d'analyse mathématique, d'estimation statistique et de prédiction qui soient adaptés aux problématiques modernes, et qui se déclinent à travers la construction d'algorithmes novateurs programmés dans des langages informatiques performants. Leur maîtrise offrira aux data scientists de cette formation une visibilité et une originalité rare et précieuse sur le marché du travail.

Public visé

Professionnels possédant un bagage mathématique solide souhaitant acquérir un véritable savoir-faire opérationnel et une très bonne maîtrise des techniques statistiques et des outils informatiques nécessaires au traitement des données.

Bac+3 en mathématiques ou niveau admis comme équivalent. La qualité de l'expérience professionnelle peut être prise en considération pour accepter des candidats ne disposant pas d'un diplôme Bac+3.

Modalités pratiques

300 heures de formation réparties sur 17 mois dans les locaux de l'Université Paris Dauphine-PSL. L'organisation des cours permet la continuité d'une activité professionnelle à temps complet.

Admission

L'admission dans la formation est prononcée après :

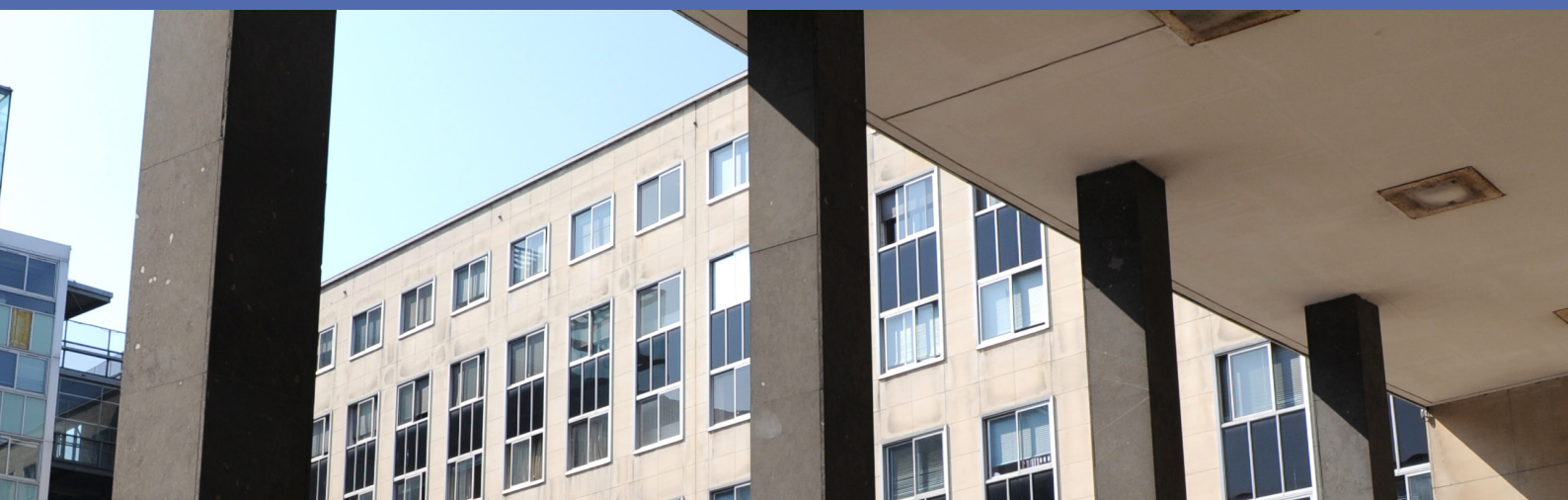
- L'examen du dossier de candidature, intégrant notamment un CV, une lettre de motivation.
- Un entretien et un test mathématique (visant à s'assurer des pré-requis nécessaires).

Corps professoral

Le programme est dirigé par Robin Ryder, maître de conférences au sein de l'Université Paris Dauphine-PSL.

L'équipe pédagogique est composée d'enseignants-chercheurs de Dauphine et d'autres universités et de professionnels dont notamment :

- Marc Hoffmann, Professeur de Mathématiques appliquées à Dauphine, Directeur de MIDO
- Vincent Rivoirard, Professeur de Mathématiques appliquées à Dauphine, Directeur du CEREMADE
- Fabrice Rossi, Professeur d'Informatique et de Mathématiques appliquées à l'Université Paris I - Panthéon Sorbonne





J'ai apprécié le fait que les formateurs ne se limitent pas à l'explication des codes informatiques mais prennent le temps d'expliquer et de démontrer les bases mathématiques de la théorie ou de la technique abordée.

Rianala Randriambololona - Promotion 1



Programme

Module 1 : Les bases de la statistique

Ce module offre une remise à niveau en statistique. Les auditeurs doivent posséder un bon niveau en mathématiques pour pouvoir se familiariser avec « la culture de la donnée » qui est prodiguée par ce module. Le concept de modélisation aléatoire est tout d'abord introduit et les techniques élémentaires de la statistique sont présentées. Le cadre du modèle linéaire gaussien permet de formuler de manière élémentaire de nombreuses situations concrètes et illustre simplement la résolution de problématiques élémentaires. Enfin, une mise en pratique informatique à l'aide du logiciel R est enseignée.

- Remise à niveau en probabilités
- Estimation classique, tests, régions de confiance
- Modèle linéaire gaussien
- Mise en pratique statistique sous le logiciel R

Module 2 : Statistique avancée

Ce module décline les outils mathématiques et statistiques qui permettent d'envisager le traitement transversal de nombreuses problématiques. L'objectif est de présenter un très large panorama des techniques modernes à la fois sur les plans méthodologiques et pratiques.

- Choix de modèles
- Modèles linéaires généralisés
- Régression non-paramétrique
- Analyse de données
- Bases de données sous SQL
- Données fonctionnelles
- Valeurs extrêmes
- Modèles de survie
- Approche bayésienne
- Algorithmes bayésiens
- Séries temporelles
- Mini projets
- Données manquantes
- R avancé
- Python

Module 3 : Apprentissage et big data

Ce module est essentiellement dévolu à la thématique du big data qui recouvre toutes les problématiques pour lesquelles les volumes des données sont tels que celles-ci ne peuvent être gérées par les approches classiques. Les techniques idoines sont celles de l'apprentissage statistique qui se situe à l'interface de l'informatique et de la statistique.

- Estimation en grandes dimensions
- Optimisation
- Apprentissage supervisé
- Graphes
- Deep learning
- Classification non-supervisée
- Ethique et protection des données
- Calcul parallèle
- Cloud computing
- Visualisation des données
- Mise en pratique marketing
- Mise en pratique actuariat
- Case studies in Finance

Modalités pédagogiques

L'enseignement repose sur 32 cours de durées variables, répartis en 3 modules de niveau progressif. Au-delà des exposés et des exercices développés au sein de chacun des enseignements, les auditeurs auront à préparer des études de cas et des exercices entre les séances.

Pour la plupart des cours, des illustrations sur jeux de données artificielles et réelles seront proposées et il sera donc important pour les auditeurs de disposer d'un ordinateur portable pour chaque séance. Les intervenants s'assureront que les auditeurs disposeront des logiciels nécessaires mis à jour. Les auditeurs auront accès au cluster informatique du CEREMADE pour les cours qui le nécessitent.

**[executive-education.dauphine.psl.eu/formations/
executive-master-diplome-universite/statistique-big-data](https://executive-education.dauphine.psl.eu/formations/executive-master-diplome-universite/statistique-big-data)**

RESPONSABLE

Robin RYDER
Maître de conférences
en mathématiques appliquées
Université Paris Dauphine-PSL

CONTACT

Nathalie PERRA
big-data@dauphine.fr
01 44 05 47 25



UNIVERSITÉ PARIS DAUPHINE-PSL
Place du Maréchal de Lattre de Tassigny - 75775 Paris cedex 16



linkedin.com/company/formation-continue