



DIGITAL

# Data Science Starter Program



# Data Science Starter Program

Avec l'explosion et la profusion des données disponibles dans un grand nombre de domaines, les nouvelles techniques de collecte et d'analyse affectent de façon profonde l'ensemble des secteurs de la vie économique, mais aussi certains secteurs de la recherche.

La formation offre une introduction aux meilleurs outils méthodologiques pour appréhender et exploiter les données et notamment les données massives (Big Data). Une immersion dans l'univers de la donnée à travers un parcours intensif portant sur les fondamentaux en sciences des données.



## Certification

«Conduire un projet de science de données» (RS 1527)



## Vous êtes

Professionnels de profil technique ou managérial (bases en statistique et/ou en programmation) qui souhaitent acquérir une bonne compréhension des techniques d'analyse de données, ainsi qu'un savoir-faire opérationnel sur le sujet.



## Objectifs & compétences acquises

- › Intégrer les enjeux techniques et économiques pour caractériser et cadrer un projet à base de science des données
- › Connaître les grandes familles de méthodes d'utilisation des données
- › Savoir expliquer les résultats obtenus lors d'une analyse de données
- › Présenter les résultats d'un projet de sciences de données dans un cadre opérationnel
- › Utiliser les résultats dans l'aide à la décision opérationnelle (dans le cadre du métier, de l'activité)
- › Concevoir et piloter l'exploitation des données
- › Superviser la réalisation d'un projet de science de données
- › Comprendre le rôle d'une équipe de data scientists



## Programme

La certification repose sur la réalisation d'un projet (professionnel ou personnel). À ce titre, vous pourrez travailler sur votre propre jeu de données et une problématique relative à votre activité. Vous profiterez des modules de formation pour affiner votre projet en échangeant avec l'équipe pédagogique. 2 experts évalueront votre travail individuel au travers d'un rapport écrit (~10 semaines après le dernier module de formation) et d'une soutenance orale en vue de vous attribuer le certificat de compétences «Conduire un projet de science de données» (RS 1527).

ADMISSION

**INTRODUCTION À LA SCIENCE DES DONNÉES**

- › Écosystème du big data • Cycle de vie d'un projet de données • Gestion de projet « data » • Architecture informatique • Questions de confidentialité

**PRÉTRAITEMENT DES DONNÉES ET VISUALISATION**

- › Exploration de données • Nettoyage de données • Normalisation • Visualisation avec R et Python • Réduction de dimensions • Sélection et création des variables

**ANALYSE DES DONNÉES ET APPRENTISSAGE AUTOMATIQUE**

- › Introduction à l'apprentissage • Apprentissage non supervisé • Apprentissage supervisé • Sélection de modèles • Évaluation des modèles

**APPRENTISSAGE AUTOMATIQUE POUR LES DONNÉES NON-TRADITIONNELLES**

- › Filtrage collaboratif • Fouilles du web, de graphes et de textes • Personnalisation, Recommandations • Publicité sur le web

**INTRODUCTION AU DEEP LEARNING**

- › Concepts de base • Réseaux convolutionnels • Réseaux récurrents • Vecteurs de nœuds

**BASES DE DONNÉES ET OUTILS BIG DATA**

- › SQL et noSQL • Calcul distribué • Cloud computing • Map Reduce et Hadoop • HIVE/PIG • Spark

**ATELIERS DONNÉES**

- › Exploration et sélection/ingénierie des variables • Application des méthodes d'apprentissage automatique • Sélection de modèles

CYCLE DE CONFÉRENCES

MÉMOIRE PROFESSIONNEL + SOUTENANCE ORALE

CERTIFICATION

**Responsables Pédagogiques**

**Erwan Le Pennec**  
Professeur à  
l'École polytechnique  
Département de  
mathématiques  
appliquées (DMAP)

**Michalis Vazirgiannis**  
Professeur à l'École  
polytechnique (DIX)  
Département d'informatique

**Certifications**

« Conduire un projet de science de données » (RS 1527)

**Candidature**

CV + lettre de motivation  
en anglais

**Durée**

20 jours  
3 sessions par an  
Présentiel et  
projet professionnel

**Langue**

Cours en anglais et en français

**Lieu**

École Polytechnique Executive  
Education, Campus Palaiseau  
*Possibilité de déclinaison sur-  
measure en intra-entreprise*

**Contact**

Sonya SAHARI  
Sonya.Sahari@polytechnique.edu  
01 69 59 66 62



[exed.polytechnique.edu](http://exed.polytechnique.edu)

