

PYTHON POUR LA DATA SCIENCE

Cette formation qualifiante a pour vocation de permettre l'acquisition de connaissances et de compétences spécifiques liées à la programmation et à l'analyse de données sous Python.

OBJECTIFS

Faire des analyses avancées en science de données. La formation est ainsi organisée essentiellement autour de la maîtrise avancée de Python.

COMPÉTENCES VISÉES

- ▶ Maîtrise des fonctions de base de Python pour travailler sur des données.
- ▶ Maîtrise des principales bibliothèques Python de science de données.

PUBLIC

- ▶ Entreprises
- ▶ Administrations nationales et collectivités territoriales
- ▶ Forces armées
- ▶ Centres de recherche

ÉQUIPE PÉDAGOGIQUE

Marc-Arthur Diaye, professeur des universités à l'université Paris 1 Panthéon-Sorbonne, directeur de la mention Économétrie-statistiques.

Chercheurs, experts et praticiens formeront le panel des intervenants.

PROGRAMME

JOUR 1

Rappel 1 (3h)

- ▶ Introduction au langage Python
- ▶ Le langage (la syntaxe, les structures), les opérateurs et les fonctions
- ▶ Présentation des données structurées et non structurées
- ▶ L'importation des données

Rappel 2 (4h)

- ▶ La préparation et le nettoyage des données (jointure, filtre, transformation, traitement des données manquantes avec pandas, numpy et scipy)
- ▶ Présentation des bibliothèques matplotlib et bokeh : Data visualisation

JOUR 2

Review jour 1 (1h)

Apprentissage statistique 1 (2h)

- ▶ Présentation des bibliothèques scikit-learn et xgboost pour l'apprentissage statistique

Apprentissage statistique 2 (4h)

- ▶ Application des modèles de Machine Learning sur des jeux de données : choix des modèles (apprentissage supervisé vs non supervisé), optimisation des modèles, choix des meilleurs modèles, View sur Ensemble model : bagging, boosting et stacking, averaging, majority vote...

JOUR 3

Apprentissage statistique 3 : View Deep Learning (3h)

- ▶ Présentation de la bibliothèque keras pour le Deep Learning
- ▶ View sur l'algorithme de classification d'images pour le Deep Learning

Exercice pratique supervisé (3h)

- ▶ Text Mining (NLP : Natural Language Process)

Présentation du projet d'évaluation (1h)

DURÉE

3 jours

TARIFS

1 470 €

CONTACT

RENSEIGNEMENTS ET INSCRIPTIONS

Pour plus de renseignements :
katerina.chatel@univ-paris1.fr

Sessions en intra-entreprise : intra-
fcps@univ-paris1.fr