



Data science : de la statistique aux big data

Clustering et réseaux de neurones appliqués en R

PERSONNES CONCERNÉES

Que ce soit via des appareils connectés, des applications web ou bureautiques, les entreprises créent quotidiennement de la donnée dans tous les domaines (biologie, banques et assurances, industrie, start-ups, pharmacologie, sciences sociales,...). L'importance de ces données s'explique par l'information qu'elles recèlent. Avoir accès à cette information et savoir l'analyser, est devenu un enjeu pour toutes sortes d'entreprises et toutes sortes de missions. Ainsi le spectre des personnes pouvant être intéressées est assez large.

PRÉ-REQUIS

Il n'y a aucun pré-requis en statistique ou en science des données.

La formation intégrant une large partie mise en application via un langage statistique, il est conseillé d'être assez habitué à l'outil informatique.

La participation à cette formation ne nécessite pas de pré-requis complémentaires.

PRÉSENTATION ET OBJECTIFS

Data Science et big data sont des méthodes récentes qui sont de plus en plus utilisées. Ces concepts ne peuvent être compris qu'en faisant le lien avec la statistique. Loin d'être seulement une valorisation de la donnée, ces méthodes s'appuient sur la théorie statistique et se relèvent être des outils puissants pour qui sait les utiliser tout en comprenant leurs limites. Cette formation de qualité universitaire permet d'acquérir des compétences théoriques et techniques, mais aussi d'être en mesure de développer un regard critique sur ces méthodes.

Cette formation permettra aux stagiaires d'acquérir de nouvelles compétences pour valoriser les données qu'ils ont à leur disposition et d'apprendre à mieux connaître, selon les cas, leur portefeuille clients, leurs références ou toute autre forme de base de données.

COMPÉTENCES À L'ISSUE DE LA FORMATION

- > maîtriser les principes de base de la programmation en R ainsi que RStudio
- > maîtriser le vocabulaire et les notions des data sciences et des big data
- > mettre en œuvre diverses analyses (méthodes de clustering, réseaux de neurones, analyses en composantes principales)
- > résumer l'information d'un jeu de données et analyser les similarités entre les individus

PROGRAMME

Big data et statistique : clarification des notions et lien entre les deux

Premiers pas en R : présentation et prise en main de R

R un outil de data-science : gestion, tri, création de données

Analyses en Composantes Principales : mise en application d'une méthode de réduction de dimensionnalité

Clustering : présentation et mise en application de 2 méthodes de classification

Réseaux de neurones : un outil d'apprentissage statistique

MÉTHODES ET RESSOURCES PÉDAGOGIQUES

La formation mélangera des moments de présentation théorique par le formateur et de la mise en pratique par l'ensemble des participants. Les méthodes seront présentées puis mises en application sur des données en utilisant le langage R. Les résultats, ainsi que leurs limites d'interprétation, seront discutés collectivement. De manière générale, le support diffusé et distribué aux stagiaires est le fil conducteur de ces trois journées mais le programme ouvre place à des discussions plus générales. Le logiciel Rstudio sera utilisé comme support pour la réalisation des scripts de code R.

RESPONSABLE SCIENTIFIQUE

M. Nicolas POULIN, Ingénieur de Recherche, UFR de Mathématique et d'Informatique

Courriel : poulin@math.unistra.fr

INTER ENTREPRISES

Durée : 3 jours (21 heures de cours en présentiel)

En 2025

Référence : DAE25-1293A

du 26 mai 2025

au 28 mai 2025

Tarif

1490 €

Lieu

UFR de Mathématique et d'Informatique

7 Rue René Descartes

67084 Strasbourg Cedex

Ce stage peut être réalisé en intra : nous consulter.

Renseignements et inscriptions

Diane ABELÉ

Tél : 03 68 85 49 30

Sauf le vendredi après-midi

abele@unistra.fr

Nature et sanction de la formation

Cette formation constitue une action d'adaptation et de développement des compétences.

Elle donne lieu à la délivrance d'une attestation de participation.

Une évaluation en fin de formation permet de mesurer la satisfaction des stagiaires ainsi que l'atteinte des objectifs de formation (connaissances, compétences, adhésion, confiance) selon les niveaux 1 et 2 du modèle d'évaluation de l'efficacité des formations Kirkpatrick.