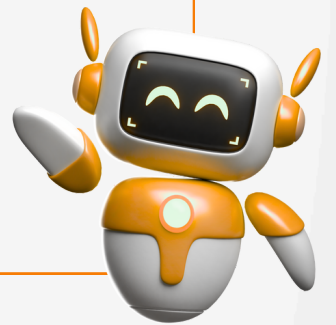




# *Plan de Formation Data Science*



## **Description du formation**



Dans l'univers du Big Data, le rôle du data scientist joue est primordial pour valoriser les données.

Cette formation Data Scientist Avancé vous apprend les modèles statistiques et les techniques d'analyse des données qui vous aideront à exploiter vos données.

Ce cours est animé par des data scientists en activité. Ils vous font bénéficier de leur retour d'expérience de modèles réels mis en place chez leurs clients et de leur expérience des compétitions de Data Science.



# Data Science



## Objectif

### Opérationnel

Savoir exploiter vos données.

## Pédagogiques

À l'issue de cette formation Data Scientist Avancé vous aurez acquis les connaissances et compétences nécessaires pour :

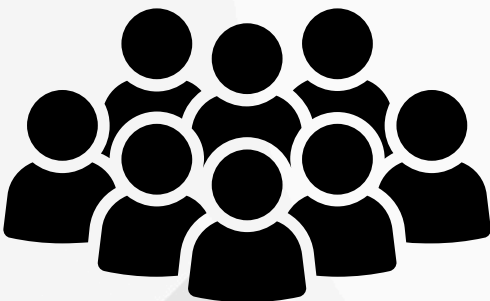
- ✓ Découvrir et utiliser la puissance prédictive des modèles ensemblistes
- ✓ Savoir effectuer un "feature engineering" performant
- ✓ Appréhender les techniques de text-mining et de deep-learning à travers d'exemples concrets
- ✓ Enrichir ses outils de data scientist

# Data Science

## A qui s'adresse cette formation ?

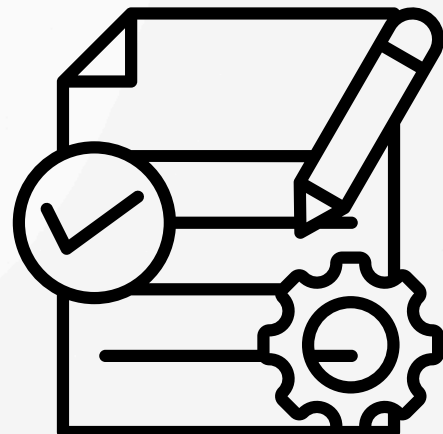
### Public

le cours s'adresse aux data scientists, analystes, statisticiens, architectes et développeurs souhaitant approfondir leurs connaissances sur la Data Science.



### Prérequis

Pour suivre cette formation, il est recommandé d'avoir des connaissances de base en programmation ou en scripting. Il faut également avoir suivi la Formation "Fondamentaux de la Data Science" (OFDS) ou avoir les connaissances équivalentes.



# Data Science

## Contenu du cours

### Rappel des fondamentaux Data Scientist

- ✓ Ecosystème du Big Data et Data Scientist
- ✓ Comment modéliser un problème de data science ?
- ✓ Les différentes familles d'algorithmes (supervisé : classification/régression, non supervisé)
- ✓ Les algorithmes classiques
- ✓ Comment évaluer la performance ?
- ✓ Sur apprentissage et compromis biais/variance

### Modèles ensemblistes

- ✓ Rappels
- ✓ Pourquoi ça fonctionne ? Raisons théoriques
- ✓ Introduction au stacking
- ✓ Architecture et promesses du stacking
- ✓ Feature weighted stacking
- ✓ Mise en application

# Data Science

## Contenu du cours

### Introduction au text mining

- ✓ Un modèle de représentation : le bag of words
- ✓ Normalisations usuelles
- ✓ Stemming, lemmatization
- ✓ Distances (Levenshtein, Hamming, Jaro-Winkler)
- ✓ Word2Vec

### Feature engineering avancé

- ✓ Normalisation
- ✓ Qu'est ce que la normalisation ?
- ✓ Quand l'utiliser ?
- ✓ Réduction de dimension (ACP, TSNE, LSA, etc.)
- ✓ Transformation et interactions entre variables
- ✓ Traitement des variables catégorielles à haute dimensionnalité
- ✓ Création de variables extraites d'arbres (Facebook Trick)

# Data Science

## Contenu du cours

### Réseaux de neurones et deep learning

---

- ✓ L'origine : le perceptron
- ✓ Les réseaux de neurones
- ✓ Deep learning
- ✓ Objectif : s'affranchir du feature engineering manuel
- ✓ Convolution
- ✓ Réseaux récurrent
- ✓ Cas concret
- ✓ Reconnaissance de chiffres

### Apprentissage semi-supervisé

---

#### ➤ Rappels et révisions

- ✓ Synthèse des points abordés les deux premiers jours
- ✓ Approfondissement des sujets sélectionnés avec l'intervenant

#### ➤ Mise en pratique

- ✓ Le dernier jour de ce stage est entièrement consacré à des mises en pratique  
Sélection et participation à une compétition
- ✓ Le formateur sélectionnera une compétition en cours
- ✓ sur Kaggle qui sera démarrée en jour 3 par l'ensemble des participants



Knowledge forges Empires

## Nos Contacts



**empire-training.tn**



**+216 55 826 628**  
**+216 74 201 616**



**16 Rue D'Athènes, Sfax**



**Contact@empire-training.tn**

