

retengr

## Pensez comme un Data Scientist FOAD

---

Durée : 4.5 jours



# Méthode pédagogique

Composée à 70% de pratique, cette formation utilise des exercices illustrés, présentés et accomplis avec le formateur, et des exercices à réaliser en autonomie.

Une journée se décompose de la façon suivante :

Matin : 2h de théorie en visioconférence, 10 à 15 minutes de présentation d'exercices en visio, 1h de TP en autonomie avec possibilité de solliciter le formateur (partage d'écran à distance pour une assistance efficace).

Après-midi : 2h de théorie en visioconférence, 10 à 15 minutes de présentation d'exercices en visio, 1h de TP en autonomie avec possibilité de solliciter le formateur (partage d'écran à distance pour une assistance efficace).

Une évaluation quotidienne de l'acquisition des connaissances de la veille est effectuée.

Une synthèse est proposée en fin de formation.

Une évaluation à chaud sera proposée au stagiaire à la fin du cours.

Un support de cours (version électronique) sera remis à chaque participant comprenant les slides sur la théorie, les exercices. L'émargement par demi-journée de présence se fera de façon numérique.

Enfin, une attestation de formation sera envoyée si le stagiaire a bien assisté à la totalité de la session.

En ce qui concerne le matériel informatique du stagiaire, il est seulement préconisé un ordinateur et une connexion internet. Nous nous chargeons du reste.

Chaque participant se verra attribuer une Machine Virtuelle qui sera exécutée dans le Cloud d'Amazon. Il disposera alors de la puissance et des outils nécessaires pour le bon déroulement de la formation. Aucune installation de la part du participant n'est requise avant la formation.



# Présentation

Plus d'un tiers des ventes du site Amazon.com est généré grâce à leur moteur de recommandation. Leurs algorithmes de Machine Learning leur a aussi révélé que les recommandations sont d'autant plus efficaces que lorsqu'elles sont faites lors d'une campagne mail plutôt qu'un affichage sur la page web.

Segmentation (clustering), prédiction, estimation, recommandation, ces méthodes d'analyse sont aujourd'hui utilisées dans l'objectif d'apporter de la valeur aux données détenues par l'entreprise.

Si l'implémentation de ces algorithmes reste de la responsabilité du mathématicien, leurs usages est maintenant rendu possible par l'apparition de librairies masquant une partie de leur complexité. A l'issue de cette formation vous aurez appris les activités du Data Scientist : Choix et préparation des données, sélection des algorithmes, apprentissage, scoring, industrialisation.



# Objectifs

- Comprendre le Machine Learning et l'Intelligence Artificielle
- Catégoriser les différentes approches : clustering, classification, régression...
- Choisir, structurer et adapter les données pertinentes pour des résultats pertinents
- Apprendre le langage python et ses bibliothèques scientifiques (scikit-learn, pandas)
- Mettre en œuvre des cas concrets (prédiction, clustering)
- Mesurer la pertinence des modèles mis en œuvre
- Traiter des gros volumes de données en parallélisant les traitements : Spark et Dask
- Déployer un modèle en production
- Introduction au Deep Learning

# Audience

Architectes, Chefs de projet, Développeurs

# Le formateur

Le formateur est un expert du domaine qui intervient sur le sujet depuis plusieurs années en formation mais aussi en conseil.

Doté d'une grande qualité d'écoute, sa pédagogie et sa compétence technique vous permettront d'acquérir les compétences sur ce sujet.

Il saura alterner entre théorie, pratique, et retours d'expérience.

# Pré-requis

La connaissance d'un langage de programmation structuré est nécessaire

# Programme

## JOUR 1 [3.5h]

### Les bases du machine learning [3.5h]

- Quelques exemples en guise d'introduction
- Généraliser : un principe fondateur du Machine Learning pour permettre la prédiction et la segmentation
- Algorithmes supervisés et non supervisés
- Appropriation du vocabulaire du Data Scientist

### Premiers pas [3.5h]

- Collecter et stocker les données
- Analyser, comprendre, nettoyer et structurer les données : Le Feature Engineering
- Apprentissage et création d'un modèle
- Evaluation du modèle
- Amélioration du modèle
- JOUR 2

### Le langage python [3.5h]

- La syntaxe du langage
- Les outils de développement : Jupyter notebook
- Les bibliothèques du data scientist : Pandas, Scikit-learn
- Analyser et comprendre les données



- Matplotlib et Seaborn : Des bibliothèques de data visualisation pour Python

## Choisir les algorithmes de machine learning [3.5h]

- Comprendre les enjeux
- Tour d'horizon des principaux algorithmes
- Classification : k-Nearest Neighbors (k-NN),
- Arbre de décision, Random Forest, XGBoost
- Régression : Régression logistique
- Clustering : K-Means, DBScan
- Les différentes méthodes de scoring
- JOUR 3

## Concepts avancés [3.5h]

- Validation croisée
- Ensemble Machine Learning : cumulez les algorithmes pour une meilleure précision
- Automatiser les manipulations de données avec un pipeline

## Traiter les données en parallèle [3.5h]

- Pourquoi paralléliser ?
- Adapter les algorithmes
- Une complexité complémentaire
- Les frameworks de distribution à disposition : Spark et Dask
- JOUR 4

## Déployer en production [3.5h]

- Intégrer un pipeline à une chaîne de déploiement automatisée (continuous delivery)
- Packager un modèle : Predictive Model Markup Language
- Créer un endpoint REST avec python Flask
- Déployer dans le cloud



## Le Deep Learning [3.5h]

- Présentation générale : les réseaux de neurones
- Les réseaux convolutionnels : Analyse d'images
- Les réseaux récurrents
- Comprendre par l'exemple : Pytorch & Keras

# Modalités et délais d'accès à la formation

Les inscriptions sont possibles jusqu'à 48 heures ouvrées avant le début de la formation, en interentreprises, dans la limite des places disponibles. Pour les formations organisées en intra entreprise, la liste des participants peut être modifiée jusqu'à 24h ouvrées avant le début de la formation.

## Accessibilité

RETENGR facilite l'accessibilité de ses formations.

Cette formation est accessible aux personnes en situation de handicap.

Si vous avez un besoin d'accès spécifique, contactez Céline BOURREIL ([celine.bourreil@retengr.com](mailto:celine.bourreil@retengr.com)) qui étudiera avec Handifiel's (notre référent handicap) votre demande et vous proposera les meilleures solutions



**Vous allez nous adorer si  
comme nous vous pensez que...**

# Une formation doit être au service de la performance du collaborateur et de l'entreprise

Ceci nécessite une quête constante d'excellence de la part de l'organisme formateur avec une adaptation systématique aux enjeux de l'entreprise, la mise à jour régulière des supports de cours et une veille technologique indispensables pour toujours être à la pointe du domaine.



# L'expertise technique est aussi importante que les qualités pédagogiques



Nos formateurs sont tous des experts de leur domaine. Mais qu'ont-ils de plus que les autres ? Nous les sélectionnons en plus pour leurs qualités de pédagogue et leurs méthodes d'enseignements. Nous plaçons les qualités pédagogiques au même niveau que l'expertise afin que nos stagiaires tirent le meilleur de leurs formations.



re'engr

## L'excellence naît de l'excellence

Beaucoup de nos clients se classent parmi les leaders de leurs industries respectives ou parmi les start-ups les plus prometteuses. Nous savons que former les collaborateurs de telles entreprises nécessite de prêter attention à chaque détail en prodiguant un accompagnement à la hauteur de l'ambition de nos stagiaires. C'est pourquoi nous savons faire des leaders d'aujourd'hui les champions de demain !





retengr

**Faire du leader  
d'aujourd'hui, le champion  
de demain**