

retengr

# Pensez comme un DataScientist

---

Durée : 4 jours – Réf : ML

Taux moyen de satisfaction en 2023 : 94%

Heures par stagiaires en 2023 : 896



- 40 rue des Filatiers 31000
- Organisme de formation : 73 31 08023 31
- SIRET : 812 673 903 00017

# Méthode pédagogique

Composée à 70% de pratique, cette formation utilise des exercices illustrés et didactiques.

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 sera remis à chaque participant comprenant les slides sur la théorie, les exercices. Une feuille d'émargement par demi-journée de présence est fournie en fin de formation ainsi qu'une attestation de fin de formation si le stagiaire a bien assisté à la totalité de la session.

Un suivi et un échange avec les participants seront mis en place quelques jours 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 bibliothèques 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, Développeurs, Chefs de projet technique

# Pré-requis

# Positionnement

La connaissance d'un langage de programmation structuré est nécessaire.

Afin de valider les compétences et les pré-requis de chaque participant, en amont de la formation, le formateur organise un entretien téléphonique. Il confirme alors que le participant a le niveau nécessaire et que le contenu répond bien à ses attentes.



- 40 rue des Filatiers 31000
- Organisme de formation : 73 31 08023 31
- SIRET : 812 673 903 00017

# 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 de comprendre le métier du Data Scientist.

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

# Programme

## Les bases du machine learning

- 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

- 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

## Le langage python

- 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



- 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

## Concepts avancés

- 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

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

## Déployer en production

- 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

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



- 40 rue des Filatiers 31000
- Organisme de formation : 73 31 08023 31
- SIRET : 812 673 903 00017

# 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 ouvrés avant le début de la formation.

## Accessibilité de la formation

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

The background is a vibrant, abstract composition of overlapping shapes and patterns. It features a large yellow shape on the left, a red shape on the right, and a purple shape at the bottom. A blue shape is visible at the top. A pattern of yellow dots is arranged in a semi-circular shape on the right side. A dashed blue line forms a loop and extends across the bottom. A light pink circle is positioned in the lower right area, overlapping a teal shape.

**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.





- 40 rue des Filatiers 31000
- Organisme de formation : 73 31 08023 31
- SIRET : 812 673 903 00017

# 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**