

retengr

La datascience en pratique

Durée : 6 jours



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.

Présentation

Le Machine Learning et en particulier le Deep Learning sont à l'origine de nombreux progrès dans différents domaines (vision par ordinateur, traitement naturel du langage, etc...) applicables à tous les secteurs d'activités.

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.

Vous connaîtrez également les principes fondamentaux du Deep Learning et vous maîtriserez les différentes architectures de réseaux de neurones, vous permettant de créer les modèles répondant à vos problématiques.



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
- Comprendre le Deep Learning
- Utiliser les frameworks de Deep Learning : TensorFlow, PyTorch et JAX
- Maîtriser les différentes architectures de réseaux de neurones : denses, convolutionnels, récurrents, génératifs
- Mettre en œuvre des cas concrets pour chaque type de réseaux
- Exécuter des calculs sur des CPUs, GPUs et TPUs
- Mesurer la pertinence des modèles mis en œuvre & visualiser l'apprentissage
- Déployer un modèle en production

Audience

Architectes, Chefs de projet, Data Analysts, Data Engineers, Data Scientists, 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 de comprendre le métier du Data Scientist.

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

Pré-requis

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

Afin de valider les compétences et les prérequis 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.

Programme

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

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

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

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

Premiers pas dans les neurones [3.5h]

- Le neurone formel
- Le perceptron
- Les fonctions d'activation
- La descente de gradient
- Démo : Playground réseau de neurones

Les Frameworks de Deep Learning [3.5h]

- TensorFlow
- PyTorch
- JAX
- Keras
- TP : Reconnaissance de chiffres (MNIST)

Les réseaux de neurones convolutionnels [3.5h]

- CNN vs le cortex visuel humain
- Couche de convolution & pooling
- Architecture d'un CNN
- Comment le réseau apprend-il ?
- Quelques architectures de référence
- Démo : Reconnaissance de chiffres avec CNN (MNIST)
- TP : Reconnaissance d'image



Les réseaux de neurones récurrents [3.5h]

- Définition RNN
- LSTM
- GRU
- Traitement automatique du langage naturel : Embeddings & Word2vec
- TP : Prédiction de séries temporelles
- TP : Analyse de sentiments

Les transformers et les LLM [3.5h]

- Attention is All You Need
- Mécanisme d'attention
- Panorama des LLM existants
- Closed model APIs vs OpenModel weights vs Fully open model
- RAG & AgenticAI

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) qui étudiera avec Handifeel's (notre référent handicap) votre demande et vous proposera les meilleures solutions.

The background is a vibrant, abstract composition. It features large, overlapping organic shapes in shades of purple, yellow, and red. On the right side, there is a pattern of small yellow dots arranged in a grid that tapers off towards the top. In the lower half, there are more shapes in light pink, teal, and purple, along with a dashed blue line that forms a loop and then extends outwards.

**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.

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**