

## Machine learning

### Public

Dirigeants d'entreprise (CEO, COO, CFO, SG, DRH...), DSI, les CDO, responsables informatiques, consultants, responsables de projets Big Data.

### Pré-requis

Posséder une culture informatique générale, des notions de probabilités et statistiques sont recommandés.

### Objectifs pédagogiques de la formation

À l'issue de la formation, le participant sera en mesure de :

- Positionner le Machine Learning dans la chaîne de traitement de la donnée
- Distinguer les compétences nécessaires ou les profils à recruter
- Identifier les clés de réussite d'un projet autour du Machine Learning
- Comprendre les concepts d'apprentissage automatique et l'évolution du Big Data vers le Machine Learning
- Appréhender les enjeux de l'utilisation du Machine Learning, incluant les bénéfices attendus et des exemples d'usage

### Modalités et pédagogie

Questionnaire d'évaluation en amont et en fin de formation -  
Programme entièrement personnalisé avec des cas pratiques et études de cas  
Attestation de fin de stage

### Durée :

2 jours - 14H

### Tarif :

Selon les besoins et en fonction du nombre de salarié

## Accessibilité :

Accessible aux personnes handicapées, merci de contacter notre référent Mme Sweenarain Joshna au 01 70 36 77 54.

## Programme

### Histoire du Machine Learning et contexte du Big Data

- Replacer à leur échelle les concepts d'Intelligence Artificielle, apprentissage automatique (machine learning) ...
- Le lien avec les mathématiques, les statistiques (inférentielles), le data mining et la data science.
- Passer de l'analyse descriptive à l'analyse prédictive puis prescriptive.
- Les applications du Machine Learning (moteurs de recherche, détection des spams, lecture des chèques).
- La typologie des algorithmes de Dominique CARDON.
- La communauté Data Science et les challenges Kaggle (ex. de Netflix).

### *Etude de cas*

Etudes d'applications concrètes du Machine Learning (moteurs de recherche, détection des spams, lecture des chèques).

### Les données à disposition : collecte et préparation

- Données structurées, semi-structurées et non structurées.
- Nature statistique des données (qualitatives ou quantitatives).
- Objets connectés (IoT) et streaming.
- Opportunités et limites de l'Open Data.
- Identification des corrélations, problème de la multicolinéarité.
- Réduction des dimensions par Analyse des Composantes Principales.
- Détection et correction des valeurs aberrantes.
- Les ETL (Extract Transform Load).
- Le Web scraping.

## Les outils du marché pour le traitement de la donnée et le Machine Learning

- Les logiciels traditionnels (SAS, SPSS, Stata...) et leur ouverture à l'Open Source.
- Choisir entre les deux leaders Open Source : Python et R.
- Plateformes Cloud (Azure, AWS, Google Cloud Platform) et solutions SaaS (IBM Watson, Dataiku).
- Nouveaux postes en entreprises : data engineer, data scientist, data analyst, etc.
- Associer les bonnes compétences à ces différents outils.
- Les API en ligne (IBM Watson, Microsoft Cortana Intelligence...).
- Les chatbots (agents conversationnels).

## Les différents types d'apprentissage en Machine Learning

- Apprentissage supervisé : répéter un exemple.
- Apprentissage non supervisé : découvrir les données.
- Online (Machine) Learning par opposition aux techniques batch.
- Reinforcement learning : optimisation d'une récompense.
- Autres types d'apprentissage (par transfert, séquentiel, actif...).
- Illustrations (moteurs de recommandation...).

## Les algorithmes du Machine Learning

- Régression linéaire simple et multiple. Limites des approches linéaires.
- Régression polynomiale (LASSO). Séries temporelles.
- Régression logistique et applications en scoring.
- Classification hiérarchique et non hiérarchique (Kmeans).
- Classification par arbres de décision ou approche Naïve Bayes.
- Random Forest (développement des arbres de décision).
- Gradient Boosting. Réseaux de neurones. Machine à support de vecteurs.
- Deep Learning : exemples et raisons du succès actuel.
- Text Mining : analyse des corpus de données textuelles.

### *Démonstration*

Démonstration des différents algorithmes de base sous R ou Python.

## Procédure d'entraînement et d'évaluation des algorithmes

- Séparation du jeu de données : entraînement, test et validation.
- Techniques de bootstrap (bagging).

- Exemple de la validation croisée.
- Définition d'une métrique de performance.
- Descente de gradient stochastique (minimisation de la métrique).
- Courbes ROC et de lift pour évaluer et comparer les algorithmes.
- Matrice de confusion : faux positifs et faux négatifs.

### *Démonstration*

Démonstration du choix du meilleur algorithme.

### **Mise en production d'un algorithme de Machine Learning**

- Description d'une plateforme Big Data.
- Principe de fonctionnement des API.
- Du développement à la mise en production.
- Stratégie de maintenance corrective et évolutive.
- Evaluation du coût de fonctionnement en production.

### *Démonstration*

Démonstration d'API de géolocalisation et d'analyse de sentiments.

### **Aspects éthiques et juridiques liés à l'Intelligence Artificielle**

- Missions de la CNIL et évolutions à venir.
- Question du droit d'accès aux données personnelles.
- Question de la propriété intellectuelle des algorithmes.
- Nouveaux rôles dans l'entreprise : Chief Data Officer et Data Protection Officer.
- Question de l'impartialité des algorithmes.
- Attention au biais de confirmation.
- Les secteurs et les métiers touchés par l'automatisation.