

# Formation Machine Learning : Introduction par la pratique

## Description de la formation Machine Learning:

La machine learning, ou apprentissage automatique, existe depuis les années 1950. Il fonctionne grâce à la théorie des probabilités et ne s'inscrit pas dans la même logique que la programmation standard qui se structure elle autour d'hypothèses (si ... alors ... sinon ...).

Concrètement il n'y a pas de règle de construction, le programme apprend statistiquement à l'aide de corpus déjà traduits et traverse donc une phase d'apprentissage durant lequel il traite les données et se calibre.

Bien entendu, cette approche requiert des données et avec l'avènement du Big Data conjugué à la démocratisation de la puissance de calcul et des outils disponibles facilement sur le Cloud, le Machine Learning devient une réalité beaucoup plus facile à mettre en œuvre qu'auparavant.

Cette **formation Machine Learning** vous met en capacité de tester une solution de Data Science, du traitement des données à l'application du modèle et à la visualisation des résultats. Vous serez ainsi en mesure d'accompagner un projet de Data Science dans sa globalité, tout en étendant votre boîte à outils personnelle.

## Objectifs de la formation Machine Learning:

### Objectif opérationnel :

Pouvoir accompagner un projet de Data Science avec une bonne compréhension technique.

## Objectifs pédagogiques :

Concrètement, à l'issue de cette formation **Machine Learning**, vous serez à même de :

- Appréhender la nature et la diversité des solutions apportées par la Data Science
- Découvrir et comprendre les modèles les plus utilisés en Data Science
- Être en mesure de tester une solution de Data Science, du traitement des données à l'application du modèle et à la visualisation des résultats.

## À qui s'adresse cette formation ?

### Public :

Ce cours Machine Learning s'adresse à toute personne ayant des bases en développement et souhaitant en acquérir en Data Science, notamment pour :

- Pouvoir travailler avec des Data Scientists et mieux comprendre leur métier, leurs problématiques et leurs besoins (développeur, architecte...)
- Pouvoir accompagner un projet Data Science avec une certaine compréhension technique
- Étendre sa boîte à outils (data miner, analyste, statisticien...)
- Envisager une reconversion professionnelle

### Prérequis :

Pour suivre cette formation Machine Learning, des connaissances de base en programmation sont nécessaires. Quelques notions en probabilités et statistiques seraient un plus.

## Contenu du cours Machine Learning:

Introduction au monde du Big Data et de la Data Science

L'apport et les problématiques du Big Data dans les entreprises

L'approche Machine Learning

Les applications phares du Machine Learning en entreprise

## Un premier exemple de modélisation : la détection de Spams

Introduction à la définition d'un problème de Machine Learning

Acquisition du vocabulaire associé

## Les différents types d'application du Machine Learning

Classification vs régression

Supervisé vs non-supervisé

L'apprentissage par renforcement

## Prise en main des outils

Découverte des notebooks Jupyter

Prise en main des données avec pandas

## Mise en pratique sur un problème de classification

Introduction à un exemple de problématique de classification

Le modèle des k plus proches voisins

L'arbre de décision

Introduction aux notions de précision et de rappel

## Mise en pratique sur un problème de régression

Introduction à un exemple de problématique de classification

Introduction au Machine Learning par la pratique

La régression linéaire

L'arbre de décision revisité

Introduction au MAE et au RMSE

## La validation des modèles : 1ère partie

Principe de l'évaluation d'un modèle

Introduction à la notion de paramètres et d'hyper paramètres

L'optimisation des hyper paramètres sur un jeu de test : méthode du Grid Search

Sous-apprentissage et sur-apprentissage

Mise en pratique sur les problèmes explorés précédemment

### Une approche non-supervisée : le clustering

Principe du clustering

Le clustering hiérarchique

La méthode du k-means

Mise en pratique

### Nettoyage des données : 1ère partie

Données numériques et données catégorielles

Les valeurs manquantes

Première stratégie : filtrage des données

Mise en pratique

### Exploration et visualisation des données

Introduction à plotly

Exploration des données

Exploration des résultats des algorithmes

Mise en pratique

### La validation des modèles : 2e partie

Le problème des petits jeux de données

La validation croisée

Mise en pratique

### Le processus de création d'un modèle

Les différentes étapes

La notion de baseline et le principe itératif

Mise en pratique

### Les méthodes ensemblistes

Le compromis biais-variance

La méthode du bagging

La méthode du boosting

Mise en pratique

### Le nettoyage des données : 2e partie

L'interpolation des valeurs manquantes : les stratégies de bases

L'interpolation des valeurs manquantes : stratégies à base de modélisation

Mise en pratique

### Le Feature Engineering

Utilisation des variables catégorielles : les techniques d'encodage

Création de nouvelles variables discriminantes

Mise en pratique

### Ouverture sur le Deep Learning

Aperçu du fonctionnement des réseaux de neurones

Les cas d'applications

Avantages et inconvénients par rapports aux modèles classiques

### Rappels et approfondissement

Résumé des points vus durant les deux premiers jours

Approfondissement de certains points à la demande



**ITgate**

Training

Your Gateway to Excellence

## Participation à une compétition

Sélection d'une compétition en cours sur kaggle ou autre

Participation par petits groupes