



# MACHINE LEARNING AVEC PYTHON

PROGRAMME DE FORMATION DIG17

## OBJECTIFS

Découvrir tous les concepts fondamentaux à maîtriser pour un Machine Learning Engineer.

## PROGRAMME

### MODULE 1 - INITIATION À PYTHON POUR LES DÉBUTANTS (3 JOURS) : MISE À NIVEAU

- Introduction
- Mooc sur la programmation en Python
- Pré-requis Python / Pandas : rattrapage accéléré
- Aide-mémoire PYTHON pour la DATA SCIENCE fourni

### MODULE 2 INTRODUCTION POUR LES MOINS DÉBUTANTS ET LES CONFIRMÉS EN PYTHON (3 JOURS)

- Introduction
- Quel est le programme de la formation Machine Learning?
- Installation Anaconda / Python
- Installation Anaconda (version texte)
- Installation bibliothèques requises dans la formation

### MODULE 3 - MACHINE LEARNING : LES FONDAMENTAUX

- Introduction aux K plus proches voisins - k Nearest Neighbors (kNN)
- Définition du problème : qu'est-ce que le Machine Learning?
- Introduction au dataset
- Les k plus proches voisins - k Nearest Neighbors
- Distance euclidienne
- Calculer la distance pour toutes les observations
- Randomiser et trier
- Fonction pour faire des prédictions

### MODULE 4 - EVALUER LA PERFORMANCE DU MODÈLE

- Tester la qualité des prédictions
- Les métriques d'erreur
- Erreur quadratique moyenne (MSE)
- Entraîner un autre modèle
- Racine carrée de l'erreur quadratique moyenne (RMSE)
- Comparaison des erreurs

### MODULE 5 - MODÈLE MULTIVARIÉ DES K PLUS PROCHES VOISINS

- Récapitulatif
- Supprimer des caractéristiques
- Gérer les valeurs manquantes
- Normaliser les colonnes
- Distance Euclidienne pour le cas multivarié
- Introduction à la bibliothèque Scikit-learn
- Entraîner un modèle et faire des prédictions en utilisant Scikit-learn
- Calculer l'erreur quadratique moyenne en utilisant Scikit-learn
- Utiliser plus de caractéristiques
- Utiliser toutes les caractéristiques

## PUBLIC CONCERNÉ

Chefs de projet et aux équipes de la MOE devant mener des projets en IA, aux membres d'une DSI.

## PRÉ-REQUIS

Avoir des notions en Python est un plus  
Les débutants en programmation python feront l'objet d'une remise à niveau.

## DURÉE

42 heures

## TARIF INTER € TARIF INTRA €

2 940,00 €

Nous consulter

## MOYENS TECHNIQUES & PÉDAGOGIQUES

Support de cours, salle et équipement informatique mis à disposition. Plateforme de visioconférence pour les cours à distance. Evaluation des besoins en amont de la formation et adaptation du programme aux attentes. Alternance d'exercices pratiques corrigés, de questionnaires.

## QUALIFICATION DES INTERVENANTS

Formateurs disposant d'une d'expérience minimum de 5 ans sur la thématique enseignée.

## MODALITÉ D'ÉVALUATION

Remise d'une attestation individuelle en fin de bilan. Cette attestation valide le niveau de progression et les acquis.

## ACCESSIBILITÉ HANDICAP

Formation ouverte aux personnes en situation de handicap. Contactez-nous pour étudier les adaptations nécessaires ou les solutions alternatives.

Délai minimum d'accès à la formation de 15 jours à partir de confirmation du financement.





## PROGRAMME (SUITE)

### MODULE 6 - OPTIMISATION HYPER PARAMÉTRIQUE

- Récapitulatif
- Optimisation hyper paramétrique
- Elargir la grille de recherche
- Visualiser les valeurs des hyper paramètres
- Varier les caractéristiques et les hyper paramètres
- Pratiquer le déroulement des opérations (workflow)

### MODULE 7 - VALIDATION CROISÉE

- Introduction
- Validation croisée Holdout
- Validation croisée des K-Fold
- Première itération
- Fonction pour entraîner des modèles
- Exécuter une validation croisée des K-Fold en utilisant Scikit-learn
- Explorer différentes valeurs de K
- Compromis Biais - Variance

### MODULE 8 - PROJET GUIDÉ 1 : PRÉDICTION DE PRIX DE VOITURES

- Introduction au dataset
- Solution - Introduction au dataset
- Nettoyage de données
- Solution - Nettoyage de données
- Modèle univarié
- Solution - Modèle univarié
- Modèle multivarié
- Solution - Modèle multivarié
- Variation des hyperparamètres
- Solution - Variation des hyperparamètres

### MODULE 9 - REGRESSION LINEAIRE APPLIQUÉE AU MACHINE LEARNING

- Modèle de Régression Linéaire
- Apprentissage basé sur une instance Vs. Apprentissage basé sur un modèle
- Introduction aux données
- Régression linéaire simple
- Méthode des moindres carrés
- Entraîner un modèle de régression linéaire avec Scikit-learn
- Faire des prédictions
- Régression linéaire multiple

### MODULE 10 - SÉLECTION DES CARACTÉRISTIQUES

- Valeurs manquantes
- Corrélation des colonnes de caractéristiques avec la colonne cible
- Heatmap de la matrice de corrélation
- Entraîner et tester le modèle
- Supprimer les caractéristiques de faible variance
- Modèle final

### MODULE 11 - GRADIENT DESCENT

- Introduction au Gradient Descent
- Gradient Descent à un seul paramètre
- Dérivée de la fonction de coût d'erreur
- Fonction Gradient Descent
- Fonction dérivée
- Comprendre le gradient descendant pour des paramètres multiples
- Gradient de la fonction coût pour plusieurs paramètres

### MODULE 12 - MOINDRES CARRÉS ORDINAIRES

- Estimation des moindres carrés ordinaires
- La fonction de coût
- Dérivée de la fonction de coût
- Gradient Descendant vs. Moindres carrés ordinaires

### MODULE 13 - TRAITEMENT, NETTOYAGE ET TRANSFORMATION DES CARACTÉRISTIQUES

- Introduction
- Caractéristiques de catégorie
- Technique pour gérer les caractéristiques de catégorie: colonnes factices
- Transformer des caractéristiques numériques impropres ou inappropriées
- Valeurs manquantes
- Imputer des valeurs manquantes

### MODULE 14 - PROJET GUIDÉ 2 : PRÉDICTION DES PRIX DE VENTE DE MAISONS

- Introduction au projet
- Solution - Introduction au projet
- Traitement des caractéristiques
- Solution - Traitement des caractéristiques
- Sélection des caractéristiques
- Solution - Sélection des caractéristiques
- Entraînement et Test
- Solution - Entraînement et Test

### Module 15 - PARTIE 3 -- MACHINE LEARNING : TECHNIQUES INTERMÉDIAIRES

- Régression Logistique
- Classification
- Introduction aux données
- Régression logistique
- Fonction logistique
- Entraîner un modèle de régression logistique
- Tracer les probabilités
- Prédiction des libellés



## PROGRAMME (SUITE)

### MODULE 16 - INTRODUCTION À L'ÉVALUATION D'UN CLASSIFICATEUR BINAIRE

- Introduction aux données
- Précision
- Résultats de la classification binaire
- Sensibilité
- Spécificité

### MODULE 17 - CLASSIFICATION MULTI-CLASSE

- Introduction aux données
- Variables Nominales
- Classification Multi-Classe
- Entraîner un modèle de régression logistique multi-classe
- Tester les modèles
- Choisir l'origine de la voiture

### MODULE 18 - OVERFITTING (SURAPPRENTISSAGE)

- Introduction aux données
- Biais et Variance
- Compromis Biais - Variance
- Modèles multivariés
- Validation croisée
- Graphique compromis erreur quadratique moyenne / variance

### MODULE 19 - LES BASES DU CLUSTERING

- Aperçu du Clustering
- Introduction aux données
- Explorer les données
- Distance entre les Sénateurs
- Clustering
- Explorer les Clusters
- Explorer les Sénateurs présents dans le mauvais Cluster
- Tracer les Clusters
- Trouver les Sénateurs radicaux

### MODULE 20 - K-MEANS CLUSTERING

- Clustering sur les joueurs de NBA
- Les meneurs de Jeu
- Points par Match
- Ratio Passe décisive - Perte de balle
- Visualiser graphiquement ces meneurs de jeu
- Clustering des joueurs
- L'algorithme Setup K-Means (itératif)
- Visualiser les centroïdes
- L'algorithme Setup K-Means (la suite)
- Etape 1 (Distance Euclidienne)
- Etape 1 (la suite)
- Visualiser les Clusters
- Etape 2
- Répéter l'étape 1
- Répéter les étapes 1 & 2
- Utilisation de la classe KMeans

### MODULE 21 - GRADIENT DESCENT

- Introduction aux données
- Modèle linéaire
- Introduction à la fonction Cost
- La fonction Cost (suite)
- Fonction Cost - Pentas
- Algorithme du Gradient Descent
- BONUS : Introduction au concept de Réseaux de neurones
- Introduction aux données
- Neurones
- Fonction Cost
- Calculer les Gradients
- Réseau de neurones à 2 couches
- Réseaux de neurones
- Fonction de coût pour un réseau de neurones multiple
- Backpropagation (ou rétropropagation)
- Séparation des données
- Prédiction sur les fleurs d'iris
- Projet guidé 3 : Prédire le marché boursier
- Le marché boursier et le dataset
- Lire les données
- Solution - Lire les données
- Générer des indicateurs
- Solution - Générer des indicateurs
- Séparer les données
- Solution - Séparer les données
- Faire des prédictions
- Solution - Faire des prédictions
- Améliorer le modèle
- Solution - Améliorer le modèle