
Programme de Formation

Machine Learning en Python : créer vos premiers modèles prédictifs

Description

Apprenez sur des cas concrets à créer, optimiser, et déployer des modèles prédictifs de Machine Learning en utilisant Python et la librairie scikit-learn.

Points Forts

Variable non renseignée



Public visé

- Développeur, data engineer, data analyst, data scientist, chercheur, ingénieur R&D, chef de projet technique, statisticien, et toute personne travaillant dans la data et sachant manipuler du code informatique



Objectifs pédagogiques

- Comprendre l'apport du Machine Learning et ses limites
- Maîtriser les principaux algorithmes
- Savoir créer et optimiser un modèle prédictif en python
- Mesurer la qualité des modèles et les performances attendues en production
- Connaître le workflow global du projet
- Savoir mettre en oeuvre les bonnes pratiques pour éviter les écueils de ce type de projet
- Pouvoir déployer un modèle, le superviser et le mettre à jour en production
- Comprendre comment appliquer le Machine Learning sur des données structurées, sur du texte, et sur des séries temporelles

Organisation

Durée : 21 heures

Mode d'organisation : Mixte

Contenu pédagogique



Description

JOUR 1

- Introduction au Machine Learning :
 - Principe général et concepts basiques
 - Exemples de cas d'usage dans différents secteurs : industrie, marketing, IoT, web, énergie...
 - Cadre d'utilisation : possibilités et limitations
 - Bien formuler la problématique : comment passer d'un problème métier à un problème Machine Learning
- Ecosystème Python :
 - Python scientifique : numpy, pandas, matplotlib, scipy
 - La librairie Scikit-learn
 - Notebook Jupyter, Anaconda



- Algorithmes de Machine Learning, première partie : les bases
- Régression linéaire et régression logistique
- K plus proches voisins : KNN
- Arbres de décision et Random Forests

JOUR 2

- Critères d'évaluation :
 - Régression : MAE, MSE, RMSLE, R^2 ...
 - Classification : accuracy, precision, recall, F1 score...
 - Procédures d'évaluation : train-test split, cross-validation, validation set
- Optimisation des hyper-paramètres :
 - Gridsearch, randomsearch
 - Soft optimisation et hard optimisation
- Méthodologie et bonnes pratiques :
 - Déroulé d'un projet de data science : une procédure itérative
 - Workflow complet du projet
 - Pipeline de transformation
 - Ecueils à éviter et comment s'en prémunir : surapprentissage (overfitting) et fuite de données (data leakage)

JOUR 3

- Data preparation et feature extraction :
 - Traitement des données aberrantes et manquantes
 - Normalisation et standardisation
 - Combinaison de features
- Mise en production :
 - Déployer un modèle en production via une API
 - Monitoring des modèles
 - Mise à jour des modèles
- Algorithmes de Machine Learning, deuxième partie : les autres catégories
 - Boosting et gradient boosting
 - Clustering
 - Détection d'anomalie
 - Réseaux de neurones et Deep Learning
- Adapter selon le type de données :
 - Comment traiter du texte
 - Comment traiter des séries temporelles

★ Prérequis

- Connaître un langage de programmation, idéalement python
- Un test de positionnement sera réalisé au préalable pour vérifier si vous disposez des compétences nécessaires pour suivre la formation.



Modalités pédagogiques

Cette formation alterne contenu théorique et mise en pratique. Elle vous permettra de comprendre en profondeur le Machine Learning : ses enjeux et ses limites, comment concevoir un use case, quels sont les principaux algorithmes et comment les utiliser, les optimiser, et évaluer la qualité des modèles prédictifs obtenus.

Vous mettrez les notions apprises en pratique sur des cas concrets réels, et maîtriserez les bonnes pratiques du domaine, les écueils à éviter absolument, le cycle de développement et de déploiement, ainsi que la supervision du modèle en production.

L'objectif est de vous donner toutes les bases nécessaires pour que vous puissiez à l'issue de la formation faire votre propre projet de Machine Learning, depuis la récupération des données jusqu'à l'exploitation de votre modèle. Vous comprendrez comment appliquer ces techniques sur des données structurées, du texte, ainsi que des séries temporelles.



Moyens et supports pédagogiques

- Exercices concrets
- Cas pratiques
- Quiz d'évaluation des connaissances



Modalités d'évaluation et de suivi

- Positionnement en amont de la formation :
 - Un quiz de consolidation des pré-requis sera administré en amont de la formation
- Suivi « pendant » :
 - Feuilles de présence
 - Exercices pratiques
 - Évaluation « fin de formation »
 - Évaluation des acquis en fin de formation
 - Formulaire d'évaluation de la formation
- Évaluation à froid :
 - Suivi post-formation : Questionnaire de satisfaction à j+60