

Python – Machine Learning

Programme (Mis à jour le 12/02/2021)

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

Critères d'évaluation

- Régression : MAE, MSE, RMSLE, R²...
- 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 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 et mise à jour des modèles

Algorithmes de Machine Learning, deuxième partie : les autres catégories

- Boosting et gradient boosting
- Clustering et détection d'anomalie
- Réseaux de neurones et Deep Learning

Référence

THDI2073

Durée

3 jours / 21 heures

Prix HT / stagiaire

2175€

Objectifs pédagogiques

- Identifier l'apport du Machine Learning et ses limites
- Créer et entraîner un modèle prédictif en Python
- Concevoir l'optimisation de la qualité des modèles et mesurer leurs performances

Niveau requis

- Garantir pratiquer un langage de programmation, idéalement Python

Public concerné

- 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

Formateur

Les formateurs intervenants pour Themanis sont qualifiés par notre Responsable Technique Olivier Astre pour les formations informatiques et bureautiques et par Didier Payen pour les formations management.

Conditions d'accès à la formation

Délai : 3 mois à 1 semaine avant le démarrage de la formation dans la limite des effectifs indiqués

Moyens pédagogiques et techniques

Salles de formation (les personnes en situation de handicap peuvent avoir des besoins spécifiques pour suivre la formation. N'hésitez pas à nous contacter pour en discuter) équipée d'un ordinateur de dernière génération par stagiaire, réseau haut débit et vidéo-projection UHD

Documents supports de formation projetés
Apports théoriques, étude de cas concrets et exercices

Mise à disposition en ligne de documents supports à la suite de la formation

Dispositif de suivi de l'exécution de l'évaluation des résultats de la formation

Feuilles d'émargement (signature électronique privilégiée)

Evaluations formatives et des acquis sous forme de questions orales et/ou écrites (QCM) et/ou mises en situation

Questionnaires de satisfaction (enquête électronique privilégiée)

Adapter selon le type de données

- Comment traiter du texte
- Comment traiter des séries temporelles