

Introduction au Machine Learning avec Scikit-learn

Référence : IA049

Durée : 2 jours (14 heures)

Certification : Aucune

CONNAISSANCES PRÉALABLE

- Aucune connaissance préalable

PROFIL DES STAGIAIRES

- Débutants en Machine Learning, data scientists juniors

OBJECTIFS

- Construire des modèles supervisés et non supervisés
- Évaluer et comparer les modèles
- Préparer les données pour le ML

CERTIFICATION PRÉPARÉE

- Aucune

MÉTHODES PÉDAGOGIQUES

- Mise à disposition d'un poste de travail par stagiaire
- Remise d'une documentation pédagogique numérique pendant le stage
- La formation est constituée d'apports théoriques interactifs, de démonstrations en direct, de travaux dirigés, d'ateliers collaboratifs, de mises en situation professionnelles, d'un projet fil rouge et d'études de cas réels
- Le suivi de cette formation donne lieu à la signature d'une feuille d'émargement

FORMATEUR

- Consultant-formateur expert en Machine Learning et plus particulièrement sur Scikit-learn

MÉTHODES D'ÉVALUATION DES ACQUIS

- Validation des prérequis envoyée en amont de la formation. En l'absence de réponse d'un ou plusieurs participants, un temps sera consacré en ouverture de session pour prendre connaissance du positionnement de chaque stagiaire sur les objectifs pédagogiques évalués
- Validation progressive des compétences via exercices pratiques, quiz techniques, corrections collectives et restitution des travaux réalisés durant le projet fil rouge
- Auto-évaluation des acquis par le stagiaire via un questionnaire
- Attestation des compétences acquises envoyée au stagiaire
- Attestation de fin de stage adressée avec la facture

ACCESSIBILITÉ DE LA FORMATION

- Si vous avez besoin d'un aménagement pour suivre votre formation dans de bonnes conditions, nous vous invitons à contacter notre référent handicap au 01 71 19 70 30 ou à l'adresse referent.handicap@edugroupe.com. Un échange en amont nous permettra d'étudier avec vous les adaptations possibles.

CONTENU DU COURS

1. Jour 1 – Fondamentaux du Machine Learning et modèles supervisés (7h)

2. Introduction au Machine Learning et à l'écosystème Scikit-learn (1h30)

- Introduction au Machine Learning
- Types d'apprentissage : supervisé, non supervisé
- Cas d'usage métiers
- Présentation de Scikit-learn
- Workflow d'un projet Machine Learning
- Environnement Python Data Science (Pandas, NumPy, Matplotlib, Scikit-learn)
- Structure d'un pipeline ML simple
- 💡 *Mises en pratique : Installation et prise en main de l'environnement ; Exploration d'un premier dataset ; Construction d'un premier workflow ML*

3. Préparation et transformation des données (2h30)

- Chargement et exploration des données
- Nettoyage des données
- Gestion des valeurs manquantes
- Encodage des variables catégorielles
- Standardisation et normalisation
- Sélection et transformation des variables
- Découpage train/test
- Introduction aux pipelines Scikit-learn.
- 💡 *Mises en pratique : Nettoyage d'un dataset réel ; Construction d'un pipeline de preprocessing ; Préparation des données pour un modèle supervisé*

4. Construction de modèles supervisés (3h)

- Régression linéaire
- Classification avec Logistic Regression
- K-Nearest Neighbors
- Arbres de décision
- Random Forest
- Entraînement et prédiction
- Paramètres principaux des modèles
- Introduction au surapprentissage
- 💡 *Mises en pratique : Création de modèles de classification ; Comparaison de plusieurs algorithmes ; Analyse des prédictions et des performances*

5. Évaluation, optimisation et Machine Learning non supervisé (7h)

6. Évaluation et amélioration des modèles (3h)

- Métriques de classification
- Métriques de régression
- Matrice de confusion
- Accuracy, précision, rappel, F1-score
- Validation croisée
- Séparation entraînement/test
- Introduction au tuning des hyperparamètres
- Grid Search avec Scikit-learn
- 💡 *Mises en pratique : Évaluation de modèles supervisés ; Mise en œuvre d'une validation croisée ; Optimisation d'un modèle via Grid Search*

7. Introduction au Machine Learning non supervisé (2h)

- Principes du clustering
- Algorithme K-Means
- Réduction de dimension
- Introduction à la PCA
- Cas d'usage métiers
- Évaluation des clusters
- 💡 *Mises en pratique : Segmentation d'un jeu de données ; Visualisation des clusters ; Analyse des résultats obtenus*

8. Atelier fil rouge : création d'un projet Machine Learning complet (2h)

- Analyse du besoin
- Préparation des données
- Sélection et entraînement des modèles
- Évaluation comparative
- Optimisation simple
- Interprétation des résultats
- 💡 *Mises en pratique : Réalisation d'un mini-projet Machine Learning ; Présentation des résultats ; Restitution collective et recommandations*