

# Introduction aux modèles de machine learning supervisé : régressions et classification

3 j (21 heures)

Ref : IAD004

## Public

Professionnels des données et analystes commerciaux cherchant à maîtriser les modèles de Machine Learning supervisé, tels que les régressions et la classification, pour optimiser les prévisions et les stratégies décisionnelles dans leur secteur d'activité

## Pré-requis

Connaissances en statistiques et en algorithmique  
Maîtrise des bases de Python et de la manipulation de données (numpy, pandas)  
Expérience en analyse de données ou en data science

## Moyens pédagogiques

Formation réalisée en présentiel ou à distance selon la formule retenue  
Nombreux exercices pratiques et mises en situation, échanges basés sur la pratique professionnelle des participants et du formateur, formation progressive en mode participatif. Vidéoprojecteur, support de cours fourni à chaque stagiaire

## Modalités de suivi et d'évaluation

Feuille de présence émargée par demi-journée par les stagiaires et le formateur  
Exercices de mise en pratique ou quiz de connaissances tout au long de la formation permettant de mesurer la progression des stagiaires  
Questionnaire d'évaluation de la satisfaction en fin de stage  
Auto-évaluation des acquis de la formation par les stagiaires  
Attestation de fin de formation

Cette formation de 3 jours offre une introduction pratique aux modèles de machine learning supervisé pour la régression et la classification. À travers des études de cas concrets, les apprenants découvriront les principaux algorithmes utilisés pour prédire une valeur numérique (régression) ou une catégorie (classification) à partir de données étiquetées. Ils apprendront à entraîner et évaluer ces modèles sur Python avec la librairie scikit-learn sur des jeux de données métiers. Une attention particulière sera portée à l'interprétation des résultats et aux techniques de feature engineering pour améliorer les performances. À l'issue de la formation, les participants seront capables de mettre en œuvre ces algorithmes sur des projets data en entreprise de façon autonome.

Avec ce programme complet, les participants acquerront les compétences nécessaires pour appliquer les modèles de machine learning supervisé dans des contextes réels et en tirer des insights business précieux.

## Objectifs

- Comprendre les principes du machine learning supervisé
- Connaître les principaux algorithmes de régression et de classification

Introduction aux modèles de machine learning supervisé : régressions et classification

- Savoir préparer des données étiquetées pour entraîner un modèle supervisé
- Être capable d'entraîner, évaluer et optimiser ces modèles sur Python avec scikit-learn
- Maîtriser les techniques de feature engineering et de sélection de variables
- Savoir interpréter les résultats et formuler des recommandations business
- Connaître les limites et les biais potentiels des modèles supervisés

## Programme détaillé

### REGRESSION ET FEATURE ENGINEERING

---

#### INTRODUCTION AU MACHINE LEARNING SUPERVISE

---

- Différences entre régression et classification
- Principaux cas d'usage (prédiction de ventes, détection de spam...)
- Étapes d'un projet de ML supervisé

#### ALGORITHMES DE REGRESSION

---

- Régression linéaire simple et multiple
- Métriques d'évaluation (RMSE, R2...)

#### FEATURE ENGINEERING POUR LA REGRESSION

---

- Scaling et normalisation des variables
- Gestion des variables catégorielles
- Création de nouvelles variables
- Sélection de variables

#### ÉTUDE DE CAS : PREDICTION DES PRIX DE L'IMMOBILIER

---

#### PRE-PROCESSING DES DONNEES ET FEATURE ENGINEERING

---

- Entraînement et évaluation de différents modèles
- Interprétation des coefficients et des erreurs

#### CLASSIFICATION ET INTERPRETABILITE

---

#### ALGORITHMES DE CLASSIFICATION

---

- Régression logistique
- Arbres de décision
- SVM
- Métriques d'évaluation (accuracy, precision, recall, f1-score...)

## **FEATURE ENGINEERING POUR LA CLASSIFICATION**

---

Encodage des variables catégorielles (one-hot, label...)

Discrétisation des variables numériques

Gestion du déséquilibre des classes

## **INTERPRETATION DES MODELES DE CLASSIFICATION**

---

Matrice de confusion

Courbe ROC et AUC

Feature importance

LIME et SHAP pour expliquer les prédictions

## **ÉTUDE DE CAS ET CONCLUSIONS**

---

### **ÉTUDE DE CAS : PREDIRE L'ACHAT D'UN CLIENT**

---

Pré-processing des données et feature engineering

Comparaison de différents classifieurs

Optimisation des hyperparamètres

Interprétation des coefficients

## **BONNES PRATIQUES ET PIEGES A EVITER**

---

Choix de la métrique d'évaluation

Gestion du surajustement (overfitting)

Dangers des variables de fuite (leakage)

Biais des données d'entraînement

---