Introduction au Machine Learning non supervisé appliqué au business : clustering et knn

2 j (14 heures) Ref : IAD002

Public

Professionnels des données et analystes commerciaux cherchant à approfondir leur compréhension du Machine Learning, pour améliorer la segmentation et les analyses prédictives dans un contexte commercial.

Pré-requis

Expérience dans l'analyse de données (data analyst, data scientist...)
Connaissances de base en statistiques et en algorithmique
Connaissance du machine learning supervisé
Pratique de la programmation, idéalement en Python

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 2 jours offre une introduction pratique aux techniques de machine learning non supervisé et à leur application en entreprise. À travers des études de cas concrets, les participants découvriront comment extraire des insights business à partir de données non labelisées en utilisant des algorithmes comme le clustering, la réduction de dimension ou la détection d'anomalies. Ils apprendront à identifier les cas d'usage pertinents, à préparer les données et à interpréter les résultats pour répondre à des problématiques métiers. Des exercices sur Python leur permettront de mettre en pratique ces concepts sur des jeux de données pour guider la prise de décision.

Avec ce programme complet, les participants auront une vision approfondie du machine learning non supervisé et seront prêts à l'appliquer dans leurs projets professionnels.

Objectifs

Comprendre les principes et les cas d'usage du machine learning non supervisé

Savoir préparer et pré-processer des données non labelisées

Maîtriser les principaux algorithmes non supervisés : clustering, réduction de dimension, détection d'anomalies

Être capable d'entraîner ces modèles sur Python avec scikit-learn
Savoir interpréter et visualiser les résultats pour en tirer des insights business
Connaître les limites et les pièges à éviter dans un projet de ML non supervisé

Programme détaillé

INTRODUCTION AU MACHINE LEARNING NON SUPERVISE

Différences entre apprentissage supervisé et non supervisé
Principaux cas d'usage (segmentation, détection de tendances, réduction de dimension...)
Panorama des algorithmes et des librairies en Python

PRE-PROCESSING DES DONNEES NON ETIQUETEES

Gestion des données manquantes
Normalisation et standardisation
Feature selection
Exercice : pré-processing d'un dataset Kaggle

CLUSTERING

Principe du partitionnement de données
Présentation des algorithmes k-means, KNN, DBScan
Évaluation de la qualité des clusters
Étude de cas : segmentation client

APPROFONDISSEMENTS ET PROJET

PROJET FIL ROUGE

Présentation du dataset et de la problématique business
Entraînement des modèles en groupes
Interprétation des résultats et recommandations business
Restitution des projets

BONNES PRATIQUES ET ECUEILS A EVITER

Choix de l'algorithme et des hyperparamètres
Interprétabilité et explicabilité des modèles
Dérives et biais des algorithmes non supervisés
Limites du ML non supervisé

CONCLUSION ET PERSPECTIVES

RECAPITULATIF DES CONCEPTS CLES ABORDES

Discussion sur les applications potentielles dans différents secteurs d'activité

Perspectives d'évolution et de développement des compétences en ML non supervisé