Formation Préparation des données & Feature Engineering

Informations

Durée : 2 jours (14h.)

Tarif* : Nous consulter

Réf : PDFE

Niveau : Facile

intra

Mise à jour le 30/12/25

*tarif valable jusqu'au 31/12/2026

Prochaines sessions

Contactez-nous pour connaitre nos futures sessions.

Pré-requis

- Connaissances de base en manipulation de données
- Notions générales en data ou analyse recommandées
- Bases de Python ou d'outils data appréciées mais non obligatoires

Objectifs

Objectifs pédagogiques :

- Comprendre le rôle de la préparation des données dans les projets data & IA
- Identifier les problèmes courants de qualité des données
- Comprendre les principes du feature engineering
- Connaître les principales techniques de transformation des données
- Appréhender les bonnes pratiques de préparation des jeux de données

Objectifs opérationnels :

- Nettoyer et préparer un jeu de données brut
- Gérer les données manquantes et aberrantes
- Transformer des variables pour les rendre exploitables
- Créer des variables pertinentes (features)
- Préparer un jeu de données prêt pour l'analyse ou l'IA

Programme

Jour 1 – Préparation et nettoyage des données

Introduction à la préparation des données

Pourquoi préparer les données ?
Place de la préparation dans un projet data & IA
Types de données (numériques, catégorielles, textuelles)
Qualité des données et impacts métier

Identification des problèmes de données

Données manquantes
Données incohérentes ou erronées
Valeurs aberrantes (outliers)
Données dupliquées

Techniques de nettoyage

Suppression ou correction des anomalies
Gestion des valeurs manquantes
Standardisation des formats
Bonnes pratiques de traçabilité

Atelier pratique

Analyse d'un jeu de données brut
Identification des problèmes de qualité
Nettoyage et préparation initiale

Jour 2 – Feature Engineering et préparation pour l'IA

Introduction au feature engineering

Qu'est-ce qu'une feature ?
Pourquoi le feature engineering est crucial
Features manuelles vs automatiques

Formation Préparation des données & Feature Engineering

Lien entre features et performance des modèles

Transformation des variables

Encodage des variables catégorielles
Normalisation et standardisation
Discrétisation des variables
Transformation de variables temporelles

Création de nouvelles features

Agrégations et calculs dérivés
Features basées sur le temps
Combinaisons de variables
Bonnes pratiques et pièges à éviter

Atelier pratique

Création de nouvelles variables à partir des données préparées
Préparation d’un jeu de données final
Validation collective des features produites