

Data science : les fondamentaux

À l'issue de cette formation, les stagiaires seront capables d'identifier les enjeux de la Data Science et de l'IA, de schématiser le cycle d'un projet Data Science, d'appliquer les meilleures pratiques en matière de nettoyage et de préparation des données avant l'analyse, et de modéliser un problème de Data Science.

Pré-requis : Bases en statistiques, langage Python.



- Modalité :**
- Distanciel en classe virtuelle
 - E-learning : à venir
 - Présentiel

Communauté: community.reconvert.net

Durée totale : 14 H (2 jours)

Modélisation en Data Science

Modèles statistiques et algorithmes de machine learning

ACP : Analyse en Composantes Principales

Régression : linéaire, multilinéaire, polynomiale

Régression logistique

Classification supervisée

Arbres de décision et forêts

KNN (K-Nearest Neighbors)

Le K-Means clustering

Algorithme de Support Vector Machines

Réseaux de neurones : introduction

TP : choix de modèles, métriques, évaluation

PLAN DETAILLE

Introduction à la Data Science

Qu'est-ce que la Data Science ?

Data Science et IA: enjeux et applications

Outils et environnements de travail

TP : cas d'utilisation (discussion)

Apprentissage automatique et Data Science

Phases d'un projet Data Science

Apprentissage supervisé

Apprentissage non supervisé

Apprentissage par renforcement

Importance de la documentation et de la communication

TP : schématiser un projet Data Science en équipes

Rappels sur le langage Python

Introduction, installation

Syntaxe : types, structures, fonctions, modules

manipuler des fichiers : csv, txt, json, excel

La programmation objet en python

Les bibliothèques phares

Manipuler: NumPy, SciPy, Pandas, Matplotlib

Python et les bases de données

Nettoyage et préparation des données

Types et qualité des données

Techniques de nettoyage des données

Gestion des valeurs manquantes

Détection des doublons

Transformation des données : normalisation, encodage

TP: nettoyer un jeu de données réel avec Python

Exploration et analyse des données

Techniques de visualisation des données

Statistiques descriptives et inférentielles

Identification des relations et des patterns

TP : explorer les données avec Pandas et Matplotlib

Data science : les fondamentaux

À l'issue de cette formation, les stagiaires seront capables d'identifier les enjeux de la Data Science et de l'IA, de schématiser le cycle d'un projet Data Science, d'appliquer les meilleures pratiques en matière de nettoyage et de préparation des données avant l'analyse, et de modéliser un problème de Data Science.

Pré-requis : Bases en statistiques, langage Python.



- Modalité :**
- Distanciel en classe virtuelle
 - E-learning : à venir
 - Présentiel

Communauté: community.reconvert.net

Durée totale : 14 H (2 jours)

PLAN DETAILLE

Techniques de la Data Science

Cross-validation
Overfitting ou surapprentissage
Biais vs variance
Feature Engineering
Data Visualisation
La malédiction de la dimension
Bonnes pratiques

Data Science et Big Data

Introduction à Spark
La bibliothèque Spark MLLib
Exemple : modéliser la régression linéaire
Particularités de Spark MLLib
Aller plus loin

Conclusion et perspectives

Récapitulatif des compétences acquises
Évolution de la Data Science et tendances futures
Ressources pour approfondir les connaissances
TP: présentation des projets réalisés par les stagiaires