

Datascience avec Spark ML

L'objectif de cette formation est de comprendre les utilisations de la datascience et l'intelligence artificielle, et de mettre en oeuvre concrètement le Machine Learning : applications de la Datascience, cycles d'un projet, algorithmes de ML, modélisation et mise en oeuvre avec Spark ML.
Pré-requis : Python.



Modalité : à distance

Formule au choix : Live Training+ ou bien
Classe virtuelle Teams ou Zoom

Durée totale : 21 H (3 jours)

Projet Data Science : exemple : #5

Expression des besoins. Exploration des données
 Visualisation des données. Nettoyage des données
 Valeurs manquantes. Réduction de la dimension, ACP
 Entraîner un modèle ML. Les métriques d'évaluation
 Overfitting ou surapprentissage
 Biais vs variance. Cross-validation
 Réapprentissage. Bonnes pratiques

TP : mise en oeuvre Spark ML : #6

Extraction de caractéristiques et prétraitement
 Réduction de dimension, ACP
 Régression linéaire, multi-linéaire
 Régression polynomiale, logistique
 Classification supervisée. Classification non supervisée
 Régression avec arbres de décision. K-means clustering
 Réseaux de neurones artificiels

PLAN DETAILLE

Introduction à la Data Science : #1

Qu'est-ce que la Data Science ?
 Statistiques, Data Science et IA
 Champs d'application de la Data Science
 Data Science : outils et algorithmes
 Les cycles d'un projet Data Science

L'apprentissage automatique : #2

Apprentissage supervisé
 Apprentissage supervisé
 Apprentissage Renforcé

Big Data et Apache Spark: introduction : #3

Introduction au big data
 Spark : présentation et écosystème
 Spark sur Databricks vs Spark sur Hadoop
 Spark Web UI. Spark Shell
 Contexte Spark. Exécuteurs Spark
 Chargement d'un fichier en Shell
 Opérations de base via Spark Shell
 Introduction à Python et PySpark
 Exemple : analyse de données, Machine Learning

Algorithmes du ML : #4

Analyse en composantes principales
 Régression linéaire, multilinéaire
 Régression polynomiale, logistique
 Classification supervisée
 Classification non supervisée
 Régression avec arbres de décision
 K-Nearest Neighbors
 K-Means clustering
 Réseaux de neurones artificiels