



# ORGANISME DE FORMATION AUX TECHNOLOGIES ET METIERS DE L'INFORMATIQUE

## Formation Spark ML

N° ACTIVITÉ : 11 92 18558 92

TÉLÉPHONE : 01 85 77 07 07

E-MAIL : inscription@hubformation.com

### Objectifs

- | Mettre en oeuvre les outils de Machine Learning sur Spark
- | Créer des modèles et les exploiter

### Public

- | Chefs de projet, architectes

### Prérequis

- | Connaissance d'un langage de programmation comme Python, Java ou Scala.

### Programme de la formation

#### Introduction

- | Rappels sur Spark : principe de fonctionnement, langages supportés.

#### DataFrames

- | Objectifs : traitement de données structurées. L'API Dataset et DataFrames
- | Optimisation des requêtes. Mise en oeuvre des Dataframes et DataSet.
- | Chargement de données, pré-traitement : standardisation, transformations non linéaires, discrétisation
- | Génération de données.

#### Traitements statistiques de base

- | Introduction aux calculs statistiques. Paramétrisation des fonctions.
- | Applications aux fermes de calculs distribués. Problématiques induites. Approximations. Précision des estimations.
- | Exemples sur Spark : calculs distribués de base : moyennes, variances, écart-type, asymétrie et aplatissement (skewness/kurtosis)

#### Machine Learning

- | Apprentissage automatique : définition, les attentes par rapport au Machine Learning
- | Les valeurs d'observation, et les variables cibles. Ingénierie des variables.
- | Les méthodes : apprentissage supervisé et non supervisé. Classification, régression.
- | Fonctionnalités : Machine Learning avec Spark, algorithmes standards, gestion de la persistance, statistiques.

#### Mise en oeuvre sur Spark

- | Mise en oeuvre avec les DataFrames.
- | Algorithmes : régression linéaire, k-moyennes, k-voisins, classification naïve bayésienne, arbres de décision, forêts aléatoires, etc ...
- | Création de jeux d'essai, entraînement et construction de modèles.
- | Prévisions à partir de données réelles.
- | Atelier : régression logistique, forêts aléatoires, k-moyennes.
- | Recommandations, recommendForAllUsers(), recommendForAllItems();

Référence	SMLDS033
Durée	2 jours (14h)
Tarif	1 610 €HT
Repas	46 €HT(en option)

### SESSIONS PROGRAMMÉES

#### PARIS

- du 24 au 25 juin 2024
- du 23 au 24 septembre 2024
- du 18 au 19 novembre 2024

[VOIR TOUTES LES DATES](#)

## Modèles

- | Chargement et enregistrement de modèles.
- | Mesure de l'efficacité des algorithmes. Courbes ROC. MulticlassClassificationEvaluator().
- | Mesures de performance. Descente de gradient.
- | Modification des hyper-paramètres.
- | Application pratique avec les courbes d'évaluations.

## Spark/GraphX

- | Gestion de graphes orientés sur Spark
- | Fourniture d'algorithmes, d'opérateurs simples pour des calculs statistiques sur les graphes
- | Atelier : exemples d'opérations sur les graphes.

## IA

- | Introduction aux réseaux de neurones.
- | Les types de couches : convolution, pooling et pertes.
- | L'approche du Deep Learning avec Spark. Deeplearning4j sur Spark.

## Méthode pédagogique

Chaque participant travaille sur un poste informatique qui lui est dédié. Un support de cours lui est remis soit en début soit en fin de cours. La théorie est complétée par des cas pratiques ou exercices corrigés et discutés avec le formateur. Le formateur projette une présentation pour animer la formation et reste disponible pour répondre à toutes les questions.

## Méthode d'évaluation

Tout au long de la formation, les exercices et mises en situation permettent de valider et contrôler les acquis du stagiaire. En fin de formation, le stagiaire complète un QCM d'auto-évaluation.

---

## Accessibilité

Les sessions de formation se déroulent sur des sites différents selon les villes ou les dates, merci de nous contacter pour vérifier l'accessibilité aux personnes à mobilité réduite.

Pour tout besoin spécifique (vue, audition...), veuillez nous contacter au 01 85 77 07 07.