Spark ML

Plan de cours N° : 1234
Durée : 2 jours (14h)

PARTICIPANTS / PRE-REQUIS

Chef de projets, architecte.

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

OBJECTIFS PEDAGOGIQUES

Savoir mettre en oeuvre les outils de Machine Learning sur Spark. Savoir créer des modèles et les exploiter.

MOYENS PEDAGOGIQUES

Réflexion de groupe et apports théoriques du formateur
Travail d'échange avec les participants sous forme de réunion-discussion
Utilisation de cas concrets issus de l'expérience professionnelle
Validation des acquis par des exercices de synthèse
Alternance entre apports théoriques et exercices pratiques (en moyenne 30 et 70%)
Remise d'un support de cours.
Assistance post-formation d'une durée de 1 an sur le contenu de la formation via notre adresse mail dédiée formateurs@atp-formation.com

MOYENS PERMETTANT LE SUIVI DE L'EXECUTION ET DES RESULTATS

Feuille de présence signée en demi-journée,
Evaluation des acquis tout au long de la formation,
Questionnaire de satisfaction,
Attestation de stage à chaque apprenant,
Positionnement préalable oral ou écrit,
Evaluation formative tout au long de la formation,
Evaluation sommative faite par le formateur ou à l'aide des certifications disponibles.

MOYENS TECHNIQUES EN PRESENTIEL

Accueil des stagiaires dans une salle dédiée à la formation, équipée d'ordinateurs, d'un vidéo projecteur d'un tableau blanc.

MOYENS TECHNIQUES DES CLASSES A DISTANCE

A l'aide d'un logiciel comme Teams, Zoom etc... un micro et éventuellement une caméra pour l'apprenant, suivez une formation en temps réel et entièrement à distance. Lors de la classe en ligne, les apprenants interagissent et communiquent entre eux et avec le formateur.
Les formations en distanciel sont organisées en Inter-Entreprise comme en Intra-Entreprise. L'accès à l'environnement d'apprentissage (support de cours, labs) ainsi qu'aux preuves de suivi et d'assiduité (émargement, évaluation) est assuré.
Les participants recevront une convocation avec lien de connexion
Pour toute question avant et pendant le parcours, une assistance technique et pédagogique est à disposition par mail et par téléphone auprès de notre équipe par téléphone au 04.76.41.14.20 ou par mail à contact@atp-formation.com

ORGANISATION

Les cours ont lieu de 9h00-12h30 13h30-17h00.

PROFIL FORMATEUR

Nos formateurs sont des experts dans leurs domaines d'intervention
Leur expérience de terrain et leurs qualités pédagogiques constituent un gage de qualité.

ACCESSIBILITE

Les personnes atteintes de handicap souhaitant suivre cette formation sont invitées à nous contacter directement, afin d'étudier ensemble les possibilités de suivre la formation.

MISE A JOUR

01/01/2021

Spark ML

Plan de cours N° : 1234
Durée : 2 jours (14h)

Introduction

- Rappels sur Spark
  - Principe de fonctionnement
  - Langages supportés

DataFrames

- Objectifs
  - Traitement de données structurées
  - 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
  - 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 persistence
  - Statistiques

Mise en oeuvre sur Spark

- Mise en oeuvre avec les DataFrames
- Algorithmes
  - Régression linéaire
  - k-moyennes
  - k-voisins
  - Classification naïve bayésienne
  - Arbres de décisions
  - Forêts aléatoires
- Création de jeux d'essai
  - Entraînement et construction de modèles
- Prévisions à partir de données réelles
- Travaux pratiques
  - Régression logistiques
  - Forêts aléatoires
  - k-moyennes
- Recommandations
  - recommendForAllUsers()
  - recommendForAllItems()

Modèles

- Chargement et enregistrement de modèles
- Mesure de l'efficacité des algorithmes
  - Courbes ROC
  - MulticlassClassificationEvaluator()
- Mesure 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
  - Opérateurs simples pour des calculs statistiques sur les graphes
- Travaux pratiques
  - 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

31, av. du Granier
38240 Meylan
Tél. 04 76 41 14 20
Fax 04 76 41 38 45

SARL au capital de 120 000€ - 390 032 050 RCS GRENOBLE
SIRET 390 032 050 00010 - APE 8559A - TVA Intracom. FR03390032050