

Plan de cours N° : 1141

Durée : 2 jours (14h)

PARTICIPANTS / PRE-REQUIS

Chef de projet, développeur, data scientist, architecte souhaitant comprendre comment organiser le traitement des données et structurer les processus de Machine Learning.

Connaissances des principes du Big Data, et des architectures techniques mises en oeuvre.

OBJECTIFS PEDAGOGIQUES

Savoir définir les étapes de préparation des données. Comprendre et mettre en oeuvre l'apprentissage automatique, les techniques de classification de données, les apports des réseaux de neurones et du Deep Learning.

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

Plan de cours N° : 1141

Durée : 2 jours (14h)

Introduction

- Zoom sur les données
 - Format, volumes, structures
 - Requêtes et attentes des utilisateurs
- Etapes de la préparation des données
- Définitions, présentation du data munging
- Le rôle du data scientist

Gouvernance des données

- Qualité des données
- Transformation de l'information en donnée
 - Qualification et enrichissement
- Sécurisation et étanchéité des lacs de données
- Flux de données et organisation dans l'entreprise
 - De la donnée maître à la donnée de travail
- Mise en oeuvre pratique des différentes phases
 - Nettoyage, enrichissement, organisation des données

Traitements statistiques de base

- Introduction aux calculs statistiques
 - Paramétrisation des fonctions
- Applications aux fermes de calculs distribués
- Problématiques induites
 - Approximations. Précision des estimations

Data Mining

- Besoin, apports et enjeux
- Extraction et organisation des classes de données
- Analyse factorielle

Machine Learning

- Apprentissage automatique
- Définition
 - Les attentes par rapport au Machine Learning
- Ingénierie des variables
- Les méthodes
 - Apprentissage supervisé et non supervisé
- Classification des données
- Algorithmes
 - Régression linéaire
 - k-moyennes, k-voisins
 - Classification naïve bayésienne
 - Arbres de décision, forêts aléatoires
- Création de jeux d'essai, entraînement et construction de modèles
- Prévisions à partir de données réelles
 - Mesure de l'efficacité des algorithmes. Courbes ROC
- Parallélisation des algorithmes
 - Choix automatique

IA

- Introduction aux réseaux de neurones
- Réseaux de neurones à convolution
 - Modèles de CNN
- Les types de couches
 - Convolution, pooling et pertes
- L'approche du Deep Learning
 - Deeplearning4j sur Spark

Les risques et écueils

- Importance de la préparation des données
- L'écueil du "surapprentissage"

Visualisation des données

- L'intérêt de la visualisation
- Outils disponibles
- Exemples de visualisation avec R et Python