LANGAGE R

Data Scientist

Plan de cours N° : 1091

Durée : 5 jours (35h)

PARTICIPANTS / PRE-REQUIS

Développeur, chef de projets proche du développement, ingénieur scientifique sachant coder.

Maîtriser l'algorithmique, avoir une appétence pour les mathématiques. La connaissance des statistiques est un plus.

OBJECTIFS PEDAGOGIQUES

Savoir mettre en place une stratégie de traitement de données avec le langage R grâce à plusieurs modèles mathématiques, le Machine Learning et le 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

LANGAGE R

Data Scientist

Plan de cours N° : 1091

Durée : 5 jours (35h)

Introduction aux Data Sciences

- Qu'est que la data science ?
- Qu'est-ce que R ?
- Qu'est-ce que le Machine Learning ?
- Apprentissage supervisé vs non supervisé
- Les statistiques
- La randomisation
- La loi normale

Introduction à R

- Installation
- Historique
- RShell
- RStudio
- Les variables
- Les fonction sd ebase
- Les fonctions algébriques
- Les conversions
- Les fichiers RData

Les instructions R

- Les blocs
- Les conditions
- Les boucles
- Les fonctions
- Les fonctions anonymes
- Le callback

Les vecteurs

- La création
- Les opérateurs de vecteur
- Les fonctions seq et c
- L'indexation
- Les facteurs

Les statistiques avec R

- Les calculs de distribution
- La loi normale
- Ecart type et variance
- Les fonctions de statistiques autres
- Les quatiles
- Les opérations ensemblistes
- La réciprocité des données

Le graphisme

- Le graphisme avec plot
- Les différents types de graphiques
- Affichage de plusieurs graphiques
- Les packages de graphique avancés

Les régressions

- Le concept de régression en statistique
- La régression linéaire
- La fonction lm()
- L'opérateur tilda ~
- Le coefficient de corrélation
- Le loss
- L'erreur quadratique moyenne
- Les régressions polynomiales
- Les régressions fonctionnelles

Les matrices

- L'algèbre linéaire en R
- Les matrices
- Le produit et la transposée
- L'inversion de matrice
- Les nombres complexes
- L'algèbre complexe
- La transformée de fourrier FFT

Les dataframes

- L'utilité de données tabulaires
- Les dataframes
- L'accès aux dataframes
- L'indexation
- Le filtrage
- Les listes

L'accès aux sources de données avec R

- L'accès aux données binaires
- L'accès aux fichiers CSV
- L'accès aux fichiers Excel
- L'accès aux fichiers JSON
- L'accès aux fichiers XML
- L'accès aux bases de données SQL
- L'accès à SQLite, PostgreSql, MySql, Oracle et SqlServer
- L'accès aux bases de données NoSql comme assandra et MongoDb

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

LANGAGE R

Data Scientist

Plan de cours N° : 1091

Durée : 5 jours (35h)

Machine Learning

- Mise en place d'une machine learning supervisé
- Qu'est-ce qu'un modèle et un dataset ?
- Qu'est-ce qu'une régression ?
- Les différents types de régression
- La régression linéaire
- Gestion du risque et des erreurs
- Quarter d'Ascombe
- Trouver le bon modèle
- La classification
- Loi normale, variance et écart type
- Apprentissage
- Mesure de la performance
- No Free Lunch
- Mise en place d'une machine learning de régression en R

Machine Learning k Nearest Neighbors

- Algorithme des k plus proches voisins (k-NN)
- Modèle de classification
- K-NN avec SciKitLearn
- Choix du meilleur k
- Sérialisation du modèle
- Variance vs Erreurs
- La normalisation des données
- La matrice de confusion
- La gestion des images et des sons

La programmation orientée objet en R

- R et l'OO
- S4
- L'opérateur %>%
- L'objet par prototypage

Le deep learning

- L'implémentation neuralnet
- Les réseaux neuronaux
- Les réseaux neuronaux supervisés
- Les réseaux neuronaux semi-supervisés
- Le modèle MLP
- La propagation
- La rétro propagation du gradient
- Dimensionnement d'un réseau
- L'implémentation deppnet
- TensorFlow
- TensorFlow et R
- Les tenseurs
- Les tenseurs complexes (image, son, matrice)
- Keras, KerasR
- Mise en place d'un réseau Keras
- Mise en place d'un réseau complexe
- Les réseaux non supervisés (from 0)

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