



Machine Learning sous Python et R

Mise à jour nov. 2023

Durée 4 jours (28 heures)

« Délai d'accès maximum 1 mois »

28 mai-31 mai
25 nov.-28 nov.
21 oct.-24 oct.
02 sept.-05 sept.

Nantes / Rennes : 4160 € HT

Brest / Le Mans : 4160 € HT

Certification : NON

OBJECTIFS PROFESSIONNELS

- Utiliser R et Python dans le cadre de modèles d'apprentissage et de fonctions algorithmiques

PARTICIPANTS

-

PRE-REQUIS

- Cette formation s'adresse au profil métier orienté Data Scientist
- Bonnes bases de statistiques et de data mining

MOYENS PEDAGOGIQUES

- Réflexion de groupe et apports théoriques du formateur
- Travail d'échange avec les participants sous forme de
- Utilisation de cas concrets issus de l'expérience professionnelle
- Validation des acquis par des questionnaires, des tests d'évaluation, des mises en situation et des jeux pédagogiques.
- Remise d'un support de cours.

MODALITES D'EVALUATION

- Feuille de présence signée en demi-journée,
- Evaluation des acquis tout au long de la formation,
- Questionnaire de satisfaction,
- 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,
- Sanction finale : Certificat de réalisation, certification éligible au RS selon l'obtention du résultat par le stagiaire

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 et de paperboard. Nous préconisons 8 personnes maximum par action de formation en présentiel

MOYENS TECHNIQUES DES CLASSES EN CAS DE FORMATION DISTANCIELLE

- A l'aide d'un logiciel comme Teams, Zoom etc... un micro et éventuellement une caméra pour l'apprenant,
- suivez une formation uniquement synchrone 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. Nous préconisons 4 personnes maximum par action de formation en classe à distance

ORGANISATION

- Les cours ont lieu de 9h à 12h30 et de 14h à 17h30.

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.

A L'ATTENTION DES PERSONNES EN SITUATION DE HANDICAP

- 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.

Programme de formation

Généralités (03h15)

- Présentation et installation du logiciel R
- GUIs et IDEs
- Prise en main et premiers pas
- Fonctions et packages
- Utilisation des aides

Rappel des concepts de base du Data Mining (03h15)

- Introduction à R - Concepts de base
- Syntaxe
- Type de données
- Import/Export de données
- Manipulation des librairies

Manipulation des données (02h45)

- Rappels des grandeurs statistiques de bases
- Calcul des grandeurs statistiques de bases avec R
- Aperçu des types de visualisation courants en Data Mining
- Fonctions et librairies de base de visualisation avec R

Data Preprocessing (02h00)

- Data Cleaning
- Data Reduction (analyse en composantes principales)
- Data Preprocessing et ACP avec R

Règles d'association (01h15)

- Rappels théoriques de l'algorithme Apriori
- Application et exploitation des règles d'association avec R

Classification et régression (08h00)

- Règles bayésiennes naïves
- Rappels théoriques
- Initiation à la librairie "e1071" de R
- Arbre de décision CART
- Rappels théoriques
- Initiation à la librairie "party" de R
- Arbre de décision RandomForest
- Rappels théoriques
- Initiation à la librairie "randomForest" de R
- Régression linéaire
- Rappels théoriques
- Initiation à la fonction lm() de R

Clustering (06h00)

- Centroid-based clustering
- Algorithme du K-means
- Fonction kmeans de R
- Hierarchical clustering
- Concepts théoriques
- Fonction hclust de R
- Density-based clustering
- Concepts théoriques
- Application dans R

Outils de support à R (01h15)

- Data preprocessing avec Dataiku
- Utilisation de R dans Dataiku