Machine Learning sous Python et R

Mise à jour : Fév. 2024

Durée : 4 jours - 28 heures

OBJECTIFS PÉDAGOGIQUES

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

PRÉREQUIS

- Bonnes bases de statistiques et de data mining

PARTICIPANTS

- Cette formation s'adresse au profil métier orienté Data Scientist

MOYENS PÉDAGOGIQUES

- 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 questionnaires, des tests d'évaluation, des mises en situation et des jeux pédagogiques
- Remise d'un support de cours

MODALITÉS D'ÉVALUATION

- 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 PRÉSENTIEL

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

MOYENS TECHNIQUES DES CLASSES À 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-Entreprises 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 reçoivent une invitation avec un lien de connexion
- Pour toute question avant et pendant le parcours, une assistance technique et pédagogique est à disposition auprès de notre équipe par téléphone au 03 25 80 08 64 ou par mail à secretariat@feep-entreprises.fr

ORGANISATION

- Les cours ont lieu de 9h à 12h30 et de 13h30 à 17h

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

ACCESSIBILITÉ

- 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. Notre organisme peut vous offrir des possibilités d'adaptation et/ou de compensations spécifiques si elles sont nécessaires à l'amélioration de vos apprentissages

Programme de formation

Généralités

- 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

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

Manipulation des données

- 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

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

Règles d'association

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

Classification et régression

- Règles bayésiennes naïves
- Rappels théoriques
- Initiation à la librairie "e1071" de R
- Arbre de décision CART
- Rappels théoriques
- Initiation à la librarie "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

- 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

- Data preprocessing avec Dataiku
- Utilisation de R dans Dataiku