

Formation : Data Mining par la pratique

Cours pratique - 3j - 21h00

Réf. DMP - Prix : 2360 € H.T.

★★★★★ 5 / 5

Le data mining consiste à découvrir des modèles, des correspondances, des motifs dans un ensemble de données numériques ou qualitatives. Cette activité s'appuie sur une trousse à outils algorithmiques qui vous seront présentés dans ce cours. La démarche data mining sera illustrée sur plusieurs projets, en utilisant R.

Objectifs pédagogiques

À l'issue de la formation, le participant sera en mesure de :

- ✓ Comprendre l'intérêt de l'approche data mining
- ✓ Traduire une problématique et savoir y répondre
- ✓ Connaître les principales méthodes de data mining
- ✓ Identifier et savoir utiliser les outils de data mining
- ✓ Poser une problématique de data mining et rechercher la méthode appropriée
- ✓ Être capable de restituer les résultats

Public concerné

Chargés d'études, responsables de projet analyse de données, responsables Infocentre, marketing ou qualité, utilisateurs et gestionnaires métiers de bases de données, futurs data scientists.

Prérequis

Connaissances basiques en statistiques ou connaissances équivalentes à celles apportées par la formation "Statistiques descriptives, introduction" (réf. UES).

Vérifiez que vous avez les prérequis nécessaires pour profiter pleinement de cette formation en faisant [ce test](#).

Modalités d'évaluation

Le formateur évalue la progression pédagogique du participant tout au long de la formation au moyen de QCM, mises en situation, travaux pratiques...

Le participant complète également un test de positionnement en amont et en aval pour valider les compétences acquises.

Programme de la formation

PARTICIPANTS

Chargés d'études, responsables de projet analyse de données, responsables Infocentre, marketing ou qualité, utilisateurs et gestionnaires métiers de bases de données, futurs data scientists.

PRÉREQUIS

Connaissances basiques en statistiques ou connaissances équivalentes à celles apportées par la formation "Statistiques descriptives, introduction" (réf. UES).

COMPÉTENCES DU FORMATEUR

Les experts qui animent la formation sont des spécialistes des matières abordées. Ils ont été validés par nos équipes pédagogiques tant sur le plan des connaissances métiers que sur celui de la pédagogie, et ce pour chaque cours qu'ils enseignent. Ils ont au minimum cinq à dix années d'expérience dans leur domaine et occupent ou ont occupé des postes à responsabilité en entreprise.

MODALITÉS D'ÉVALUATION

Le formateur évalue la progression pédagogique du participant tout au long de la formation au moyen de QCM, mises en situation, travaux pratiques...

Le participant complète également un test de positionnement en amont et en aval pour valider les compétences acquises.

1 Le projet de data mining

- La problématique du data scientist : de la donnée à l'information.
- Le vocabulaire et les concepts.
- L'exploration descriptive du jeu de données.
- Les métadonnées pour le suivi du projet de data mining.
- Rappels sur le logiciel R.

Travaux pratiques

Utilisation de R. Caractérisation descriptive, définition et saisie des métadonnées d'un jeu de données.

2 Les techniques de l'approche data mining

- Méthode basée sur la classification : identification de groupes d'individus statistiques.
- Méthode par association : mise en évidence d'une cause et d'une conséquence.
- Méthode de l'estimation : complément d'un effectif ou d'une fréquence d'un jeu de données.
- Apports du data mining pour le traitement des grandes volumétries de données.
- Méthode de segmentation : définition de critères, extension de la méthode de classification et principe des k-means.
- Méthode de prévision : importance de la temporalité et des hypothèses.

Travaux pratiques

Appréhender les différentes méthodes en fonction des besoins exprimés.

3 Les outils statistiques

- Méthodes descriptives : corrélation, classification, réseaux de Kohonen, règles d'association.
- Méthodes prédictives : régression, arbres de décision, réseaux de neurones, les K plus proches voisins.
- Mise en œuvre de classification par k-means et CAH (Classification Ascendante Hiérarchique).
- Principe des méthodes supervisées.

Travaux pratiques

Mise en pratique des différentes méthodes sous R.

4 La visualisation des données

- Les objectifs de la visualisation des données.
- Les différents types de représentations de données quantitatives.
- Concevoir des tableaux de bord.

Travaux pratiques

Création d'un tableau de bord avec R en utilisant des données quantitatives. Représentation de données quantitatives et de données qualitatives avec R.

MOYENS PÉDAGOGIQUES ET TECHNIQUES

- Les moyens pédagogiques et les méthodes d'enseignement utilisés sont principalement : aides audiovisuelles, documentation et support de cours, exercices pratiques d'application et corrigés des exercices pour les stages pratiques, études de cas ou présentation de cas réels pour les séminaires de formation.
- À l'issue de chaque stage ou séminaire, ORSYS fournit aux participants un questionnaire d'évaluation du cours qui est ensuite analysé par nos équipes pédagogiques.
- Une feuille d'émargement par demi-journée de présence est fournie en fin de formation ainsi qu'une attestation de fin de formation si le stagiaire a bien assisté à la totalité de la session.

MODALITÉS ET DÉLAIS D'ACCÈS

L'inscription doit être finalisée 24 heures avant le début de la formation.

ACCESSIBILITÉ AUX PERSONNES HANDICAPÉES

Pour toute question ou besoin relatif à l'accessibilité, vous pouvez joindre notre équipe PSH par e-mail à l'adresse psh-accueil@orsys.fr.

5 Analyse de données qualitatives et textuelles

- Spécificités de la problématique et les alternatives (analyse factorielle des correspondances, table de contingence).
- Présentation d'instanciation, de pattern, de vecteur et d'heuristique.
- Comment utiliser un espace de vecteurs, d'indexation, de scoring.
- Différents types de transformations et traitement d'un document textuel.

Travaux pratiques

Réalisation d'un traitement de données qualitatives et de données textuelles sous R.

Parcours certifiants associés

Pour aller plus loin et renforcer votre employabilité, découvrez les parcours certifiants qui contiennent cette formation :

- [Parcours certifiant Construire et implémenter des modèles de big data et d'IA - Réf. ZBG](#)

Solutions de financement

Plusieurs solutions existent pour financer votre formation et dépendent de votre situation professionnelle.

Découvrez-les sur notre page [Comment financer sa formation](#) ou [contactez votre conseiller formation](#).

Horaires

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

Les participants sont accueillis à partir de 8h45. Les pauses et déjeuners sont offerts.

Pour les formations de 4 ou 5 jours, quelle que soit la modalité, les sessions se terminent à 16h le dernier jour.

Dates et lieux

CLASSE À DISTANCE

2026 : 4 mars, 8 juin, 19 oct.

PARIS LA DÉFENSE

2026 : 4 mars, 8 juin, 19 oct.