IBM : Introduction à IBM SPSS Modeler et Data Mining

Date et durée

Code formation : OA004
Durée : 2 jours
Nombre d'heures : 14 heures

Description

Introduction à IBM SPSS Modeler et Data Mining est un cours de base qui donne un aperçu de l'exploration de données et les bases de l'utilisation d'IBM SPSS Modeler.
Les principes et la pratique en lien avec l'exploration de données sont démontrés grâce à la méthodologie CRISP-DM

Objectifs

Le cours offre une formation de base sur la façon de

- lire, explorer et manipuler des données avec IBM SPSS Modeler,
- puis créer et utiliser des modèles de réussite.

Pré-requis

Pour suivre ce cours, il faut juste avoir des connaissances générales en informatique. Il n'est pas necéssaire d'avoir une expérience en statistiques.
Il est recommandé d'avoir une connaissance de l'organisation des données de l'entreprise ainsi que des enjeux stratégiques relatifs à l'utilisation de Data Mining.

Public

Ce cours de base est destiné à:

- Tout utilisateur d'IBM SPSS Modeler sans expérience ou peu expérimenté.
- Toute personne, avec ou sans expérience, désirant maîtriser le data Mining
- Toute personne qui envisage l'achat d'outil IBM SPSS Modeler

Cette formation s'adresse aux profils suivants

Contrôleur de gestion

Programme

Introduction au Data Mining

- Expliquer les étapes du processus de modélisation CRISP-DM.
- Décrire des projets de Data Minig qui ont réussis et expliquer les raisons qui peuvent mener à l'échec d'un projet.

- Décrire les compétences et les pré-requis nécessaires pour le Data Mining

Utilisation de flux

- Décrire les différentes zones de l'interface utilisateur Modeler.
- Travailler avec des « nœuds » et « Super-nœuds ».
- Ouvrir, exécuter et enregistrer un flux.
- Accéder à la fonction d'aide au sein du Modeler.

Présentation Data Mining

- Expliquer les principaux concepts de base utilisés dans le Data Mining .
- Construire, évaluer et déployer un modèle.
- Utiliser des nœuds Sort et Filter.

La collecte des données initiales

- Expliquer les concepts de "structure de données", "enregistrements", "champs", "unité d'analyse", "stockage".
- Lire des données à partir de plusieurs formats de fichier et les exporter dans plusieurs formats

Compréhension de Données

- Examiner les distributions de champs catégoriels et continus.
- Expliquer les façons les plus courantes de traitement des données manquantes.
- Validations des données par Modeler.

Configuration de l'unité d'analyse

- Suppression des enregistrements doubles
- Agrégation des données : Elaboration des critères d’agrégation.

Intégration de données

- Ajouter des enregistrements à partir de plusieurs bases de données en un seul ensemble de données.
- Ajouter des champs de plusieurs ensembles de données dans un ensemble de données.
- Utiliser l'échantillonnage à des fins de test.

Dérivation et remplissage des champs

- Utiliser CLEM pour transformer les données.
- Utiliser le Derive node pour créer un nouveau champ.
- Utiliser le Reclassify node.
- Utiliser le Reorder Node pour réorganiser les champs.

Recherche de relations

- Examiner la relation entre deux champs catégoriels.
- Examiner la relation entre deux champs continus.
- Examiner la relation entre un champ continu et un champ catégoriel.

Introduction à la modélisation

- Objectifs de modélisation
- Initiation à la classification
- Introduction à la segmentation