EduGroupe
Accompagner pour réussir

Dataiku DSS

Référence : **PYDS034** Durée : **2 jours** Certification : **Aucune**

CONNAISSANCES PREALABLES

• Connaissance des principes de la manipulation de données et du machine learning.

PROFIL DES STAGIAIRES

• Chefs de projet,développeurs, data-scientists, utilisateurs de Dataiku DSS.

OBJECTIFS

• Savoir installer, configurer, Dataiku DSS, l'utiliser depuis l'interface web ou des API.

CERTIFICATION PREPAREE

Aucune

METHODES PEDAGOGIQUES

- Mise à disposition d'un poste de travail par stagiaire
- Remise d'une documentation pédagogique numérique pendant le stage
- La formation est constituée d'apports théoriques, d'exercices pratiques, de réflexions et de retours d'expérience
- Le suivi de cette formation donne lieu à la signature d'une feuille d'émargement

FORMATEUR

Consultant-Formateur expert Bigdata

METHODE D'EVALUATION DES ACQUIS

- Auto-évaluation des acquis par le stagiaire via un questionnaire
- Attestation des compétences acquises envoyée au stagiaire
- Attestation de fin de stage adressée avec la facture

CONTENU DU COURS

Présentation, concepts DSS

- Fonctionnalités : Mise à disposition des méthodes et outils de data-sciences à partir d'une interface graphique ou de langages de requêtage ou de programmation (Python, SQL, R)
- Notions de projet, data, dataset, recipes
- Premiers pas avec Dataiku DSS : exemples

Connexion aux données

- Connecteurs disponibles, formats de fichiers, formats spécifiques hadoop/Spark
- Chargement des fichiers, emplacement des fichiers
- Système de fichiers, HDFS, Amazon S3, stockage Google Cloud, etc.
- FTP, connexion ssh, connexion aux bases de données SQL
- Autres stockages : Cassandra, MongoDB, Elasticsearch, ...

Préparation des données

- Scripts de préparation, échantillonnage, exécution, les différents moteurs d'exécution: DSS, Spark, base de données SQL, Spark

Graphiques et statistiques

- Présentation des types de graphiques disponibles et configuration
- Echantillonnage et exécution, graphiques de base, tables, nuages de points, visualisation cartographique, etc.
- Réalisation d'exercices pratiques
- Fonctionnalités statistiques disponibles, démonstrations

Machine learning

- Fonctionnalités disponibles : apprentissage supervisé et non supervisé, algorithmes utilisés, scoring, deep learning, optimisation de modèles, ...

Flow/Recipes

- Interface graphique de gestion des cycles de traitement,des datasets, des traitements, paramètrage de l'exécution en parallèle, des données géographiques, export en format pdf ou images, etc.
- Travaux pratiques

Interfaces de programmation

- Présentation des différentes interfaces : SQL, Python, R
- Exemples d'interrogations SQL et Python
- Interfaçage avec Spark

*Notre **référent handicap** se tient à votre disposition au 01.71.19.70.30 ou par mail à referent.handicap@edugroupe.com pour recueillir vos éventuels besoins d'aménagements, afin de vous offrir la meilleure expérience possible.*