

Formation Dataiku et MLflow

D'origine française, Dataiku s'est imposée comme l'une des références mondiales parmi les outils de Data Science et Machine Learning. DSS, pour Data Science Studio, compte pas moins de 90 fonctionnalités et permet à ses utilisateurs de déployer des modèles d'apprentissage automatique très simplement et de manière collaborative. La plateforme prend finalement en charge l'industrialisation de l'ensemble de la chaîne de traitement de la donnée : collecte, préparation, optimisation... A l'issue de cette formation Dataiku, les apprenants sont en mesure d'utiliser la plateforme d'intelligence artificielle afin d'exploiter leurs ensembles de données.

Durée

2 jours

Objectifs pédagogiques

- Prendre en main la plateforme Dataiku
- Comprendre les différents modèles d'apprentissage
- Modéliser un problème pratique sous forme abstraite via Dataiku
- Identifier les méthodes d'apprentissage pertinentes pour résoudre un problème
- Appliquer et évaluer les méthodes identifiées sur un problème
- Utiliser MLflow

Public

Data scientists, ingénieurs, analystes,
chefs de projet, développeurs...

Prérequis

Connaissances de base en statistiques
ainsi qu'en programmation Python.

Programme de formation

Introduction à la formation Dataiku

Présentation générale de la formation Dataiku
Tour de table, présentation des participants,
des contextes professionnels...
Vue d'ensemble de la plateforme Dataiku
Les fonctionnalités Dataiku : nettoyage de données,
création de dashboards, deep learning...
Dataiku DSS vs Dataiku Online
Positionnement de Dataiku dans la chaîne de
traitement de la data

L'interface Dataiku DSS (Data Science Studio)

Installation et connexion
Page d'accueil
Collaboration
Importer des données
Schéma des données
Analyser les données
Graphiques
Jobs

Manipulation de données

Les Recipes (visual, code, plugin...)

Transformer les données avec le recipe Prepare
Formula
Formatter des dates
Group By
Jointures : inner, left, right, full
Mise en forme du flow
Compilation
Le Lab
Notebooks

Visualisations

Les visualisations de l'onglet Statistics
L'interface worksheet
Visualisations sous DSS
Cartographies
Dashboards

Machine Learning : les possibilités offertes par Dataiku

Les différents types d'algorithmes d'apprentissage : supervisé, non supervisé, par renforcement...

Apprentissage supervisé : régression, régression linéaire simple et classification (arbres de décisions, SVM linéaire)

Apprentissage non supervisé : segmentation ou clustering, analyse factorielle et association

Préparer un jeu de données pour le Machine Learning (feature selection, feature engineering...)

AutoML vs mode Expert

Amélioration et personnalisation d'un modèle avec MLFlow

ML Tools

Le projet MLLib : MLflow Tracking, MLflow Projects et MLflow Models

Installation et prise en main

Moyens et méthodes pédagogiques

- La formation alterne entre présentations des concepts théoriques et mises en application à travers d'ateliers et exercices pratiques.
- Les participants bénéficient des retours d'expérience terrains du formateur ou de la formatrice
- Un support de cours numérique est fourni aux stagiaires

Modalités d'évaluation

- **En amont de la session de formation**, un questionnaire d'auto-positionnement est remis aux participants, afin qu'ils situent leurs connaissances et compétences déjà acquises par rapport au thème de la formation.
- **En cours de formation**, l'évaluation se fait sous forme d'ateliers, exercices et travaux pratiques de validation, de retour d'observation et/ou de partage d'expérience.
- **En fin de session**, le formateur évalue les compétences et connaissances acquises par les apprenants grâce à un questionnaire reprenant les mêmes éléments que l'auto-positionnement, permettant ainsi une analyse détaillée de leur progression.