

Updated 05/28/2024

Sign up

Dataiku Advanced Training

2 days (14 hours)

Presentation

Master your entire data workflow with our advanced Dataiku training. Whether you're a Data Engineer, Data Analyst or Data Scientist, our advanced training program will enable you to design complex, scalable data pipelines.

We'll explore advanced data pipelines. You'll learn how to create, optimize and automate them, based on complex scenarios and precise rules. This first step is essential for anyone wishing to certify as an advanced designer on Dataiku.

You'll discover how to create, evaluate and refine models directly in the Dataiku environment. From applying your models to understanding them in depth using the tools of responsible AI.

Finally, whether you want to integrate your own code into your projects, manage that code efficiently, or even create plugins for your colleagues, this training will give you the keys to becoming a certified Dataiku developer.

For this course, we'll be using the latest version of Dataiku: [DSS 12](#).

Objectives

- Master the creation and optimization of ML models with Dataiku's Visual ML for efficient solutions
- Use Dataiku's Responsible AI for high-performance, transparent and ethical models
- Build efficient data pipelines with Dataiku for in-depth analysis and project optimization
- Automate data pipelines and integrate code into Dataiku for customization and advanced adaptability
- Deploy and monitor your projects in production with Dataiku for a smooth transition and long-term success

Target audience

- Data analysts
- Data scientists
- Data engineers
- SAS Analyst

Prerequisites

- You have a fundamental understanding of the Dataiku environment and basic functionality, ideally acquired through practical experience or our initial Dataiku training.
- You are familiar with the concepts of data science and data analysis, and you have an understanding the basic principles of machine learning
- You have already coded at least in Python and/or R, essential for taking advantage of Dataiku's scripts and notebooks, customizing analyses, and developing models.
- It is preferable to have taken our [Dataiku training course](#)

Software requirements

Licensing Dataiku.

Our Dataiku Advanced training program

Day 1

Introduction

- Introducing Dataiku's advanced features
- Overview of the tool's capabilities

Dataiku's advanced visual tools

- Advanced visual recipes
- Using variables
- Data Pipelines
- Automation
- Collaboration
- Quiz
- Summary

Introduction Machine Learning

- Concepts
- Predictive modeling
- Prediction: Regression and classification
- Clustering
- Quiz
- Summary

Analyze your data with Interactive Stats

- Interface
- Univariate and bivariate analysis
- Curve and distribution fitting
- Correlation matrix
- Principal component analysis
- Statistical tests
- Quiz
- Summary

Machine Learning

- Setting up a model
- Model adjustment
- Understanding model predictions
- Quiz
- Summary

Scoring

- Deploying your model
- Scoring Data
- Evaluating your model
- Quiz
- Summary

Advanced Machine Learning

- Model development strategies
- Model diagnosis
- Actionability
- Quiz
- Summary

Stratified or partitioned models

- Setting up a stratified model

- Quiz
- Summary

TALN - Visually

- Introduction to TALN
- Text data preparation
- Text feature management for ML
- Quiz
- Summary

Time series analysis and forecasting

- Time series fundamentals
- Time series analysis
- Time series forecasting
- Quiz
- Summary

Temporal data preparation

- Resampling recipe
- Interval extraction recipe
- Windowing recipe
- Exterma extraction recipe
- Quiz
- Summary

Day 2

Code in Dataiku

- Code notebooks
- Code recipes
- Code environments
- External IDE integration
- Quiz
- Summary

Share your code

- Introduction
- Libraries
- Import from Git
- Code samples

- Practice
- Quiz
- Summary

Customized ML templates

- Custom pre-processing in visual ML
- Custom modeling in visual ML
- Quiz
- Summary

Variables for coders

- Variable definition
- Using variables in a code recipe
- Changing variable values
- Practice
- Quiz
- Summary

Visualization

- Webapps
- Static insights
- Quiz
- Summary

Managed folders

- How to use Managed folders
- Practice
- Quiz
- Summary

Dataiku APIs

- Introduction
- The dataiku package and the public API
- Practice
- Quiz
- Summary

Deployment and monitoring

- Dataiku Govern
- Model follow-up
- Batch deployment
- Real-time API
- Quiz
- Summary

Advanced Designer Certificate exam (Optional) ML

Practitioner Certificate exam (Optional) Developer

Certificate exam (Optional)

Companies concerned

This training course is aimed at both individuals and companies, large or small, wishing to train their teams in a new advanced computer technology, or to acquire specific business knowledge or modern methods.

Positioning on entry to training

Positioning at the start of training complies with Qualiopi quality criteria. As soon as registration is finalized, the learner receives a self-assessment questionnaire which enables us to assess his or her estimated level of proficiency in different types of technology, as well as his or her expectations and personal objectives for the training to come, within the limits imposed by the selected format. This questionnaire also enables us to anticipate any connection or security difficulties within the company (intra-company or virtual classroom) which could be problematic for the follow-up and smooth running of the training session.

Teaching methods

Practical course: 60% Practical, 40% Theory. Training material distributed in digital format to all participants.

Organization

The course alternates theoretical input from the trainer, supported by examples, with brainstorming sessions and group work.

Validation

At the end of the session, a multiple-choice questionnaire verifies the correct acquisition of skills.

Sanction

A certificate will be issued to each trainee who completes the course.

