TRAINING
Scalable Machine Learning.

Scalable Machine Learning on Databricks

The future of data is AI and Machine Learning. Yet most companies still face challenges when building machine learning models. This is due to a myriad of reasons, ever-increasing dataset sizes, heavier computation cycles for state-of-the-art models, siloed analytics teams (and data sources) and many more. This is where Databricks helps immensely, providing a common data platform for analysts, engineers, developers and data scientists to collaborate, leveraging the efficiency of Apache Spark for distributed computing.

Why Databricks?

Databricks is an open and unified data analytics platform for data engineering, data science, machine learning, and analytics that is particularly useful for deploying advanced Artificial Intelligence (AI) and Machine Learning (ML) projects.

What is this course all about?

This two-day course is for those intending to use Databricks to build machine learning models to solve real-world business use cases. You will learn how to build and tune ML models with SparkML while leveraging MLOps best practices. If you work in industries such as Financial Services, Retail, Healthcare, Media & Entertainment or any other industries and want to learn how to build machine learning models at scale, then this course is perfect for you.

Who is this course aimed at?

This course is aimed at data professionals with limited knowledge of machine learning or Databricks. It's designed to provide the tools, to allow ML enthusiasts to produce enterprise-level machine learning models whilst utilising the power of Apache Spark, to tackle enormous data manipulation issues. You will leave this course with all the skills you need to get started on your machine learning journey, from data preparation right through model training and inference.

AGENDA
DAY 1

Introduction to Data Science and Machine Learning

Supervised and Unsupervised learning

Introduction to Regression

Introduction to Classification

Introduction to Clustering

Introduction to Databricks

Apache Spark

Delta Lake

Feature Engineering and Selection

Handling and Imputing Missing Values

Feature Scaling and Encoding

Dimensionality Reduction

Introduction to Feature Stores

DAY 2

Get started with Databricks and machine learning

Introduction to MLFlow

MLFlow Pipelines

Model selection and hyperparameter tuning

Hyperparameter tuning with hyperopt

Applied Regression

Overfitting and Underfitting

Regression Evaluation Metrics

Applied Supervised Learning using Tree-based Models

Introduction to Ensemble Modeling

Oversampling and Undersampling Classes

A Review of Classification Evaluation Metrics

Interpreting models

SHAP

Feature Importance

WHAT WILL YOU LEARN?

DAY 1

✓ The basic supervised and unsupervised machine learning models

✓ How to build real-world machine learning models with Databricks

✓ Learn all about the techniques for building a robust machine learning model

✓ The latest craze in machine learning, feature stores – a centralised approach to tracking model features!

DAY 2

✓ Advanced methods for tracking models – including the newly released MLFlow pipelines!

✓ Key considerations for scoring your models

✓ How to turn a black box model into something explainable

✓ Further model improvement techniques

About advancing analytics

We are a consultancy based in the UK that helps global clients meet their Advanced Analytics objectives. We focus on Microsoft technologies and specialise in solving problems using Microsoft Azure and Databricks. By championing modern technologies, we are helping move the industry forward, either by applying deep data engineering techniques or innovative data science solutions, we combine the best of both worlds for our customers.

Judge for Yourselves

At Advancing Analytics, we think the best way to show you that we know what we are talking about, is to talk about it. You can find our **experts** at events or sharing their knowledge on **YouTube** and the **Totally Skewed** podcast.

We are known in the industry for **thought-leadership** and **bleeding-edge delivery**, we work directly with the Microsoft engineering teams, to bring you long term insight. One of our values is investing in our community, we share our thoughts and finding on the **Advancing Analytics** blog. Why not take a look?

> ❝❝ The sessions were engaging, the content was easy to understand and pitched at the right level.

databricks partner

MVP Microsoft® Most Valuable Professional

Microsoft Partner
Gold Cloud Platform
Gold Data Analytics
Gold Data Platform
Gold Application Integration
Silver Project and Portfolio Management

CERTIFIED CARBON NEUTRAL® company CarbonNeutral.com

BEACONS

FIND OUT MORE BY VISITING OUR WEBSITE AT
ADVANCING ANALYTICS.co.uk