

# AI & ML development with Databricks

The training covers key topics related to the development and deployment of AI and machine learning solutions within the Databricks environment. Participants will learn practical aspects of data preparation, building ML pipelines, training and deploying models, as well as monitoring their performance. The programme is based on the latest trends and tools used in the field of data science and MLOps



## Training recipients

This training is intended for:

- Data scientists and ML engineers who wish to learn effective data preparation and model building in Databricks.
- Data engineers and DataOps professionals responsible for preparing and managing datasets for machine learning.
- MLOps specialists deploying and monitoring ML models in cloud environments.
- Teams designing ML pipelines and individuals preparing for the Databricks Certified Machine Learning Professional certification.



## Benefits

- Comprehensive knowledge of ML pipelines – You will learn how to build modular pipelines for data preparation and model training.
- Practical skills in working with Databricks – You will become familiar with tools such as PySpark, MLflow, Feature Store, and Unity Catalog.
- Integration and automation of ML processes – You will find out how to log metrics, version models, and manage features in the Feature Store.
- Advanced data preparation – You will master techniques such as EDA, imputation, encoding,

standardisation, and feature engineering.

- Model deployment and monitoring – You will learn how to deploy models into production environments and monitor their performance.



## Training program

### 1. Building ML models

- Introduction to the process of building machine learning models in the Databricks environment.

### 2. Principles of machine learning

- Discussion of key concepts and principles of machine learning, types of algorithms, model selection.

### 3. Training machine learning models in Databricks

- Practical exercises in training ML models using Databricks and PySpark.

### 4. ML model lifecycle

- Stages of working on a model: data preparation, feature engineering, training, validation, deployment.

### 5. Monitoring and logging experiments and models using MLflow

- Managing the ML model lifecycle with MLflow: experiment logging, model versioning, metric tracking.

### 6. Model performance evaluation

- Methods of assessing model quality: metrics, cross-validation, interpretation of results.

### 7. Hyperparameter tuning

- Optimising model hyperparameters using Databricks tools.

### 8. Using AutoML in Databricks

- Automating the process of building ML models with AutoML available in Databricks.

### 9. Training deep learning models in Databricks

- Introduction to training deep learning models in the Databricks environment.



## Expected preparation of the participant

- Familiarity with basic concepts and services related to AI and ML.
- Knowledge of SQL and/or Python.
- Experience working with data (e.g., completion of a data engineering module or equivalent knowledge).
- Completion of the “Fundamentals of Databricks” training is recommended.



## Training Includes

access to Altkom Akademia student portal

- Lecture (40%)
- Workshops (30%)
- Exercises (30%)

Main teaching tools: Databricks, PySpark, MLflow, Feature Store, Unity Catalog, PowerPoint presentations, lab environments, Microsoft Learn resources.



## Language

- **Training:** English
- **Materials:** English

## Examination method

The training does not conclude with an exam. However, it prepares participants for the Databricks Certified Machine Learning Professional exam.

Exam description: <https://www.databricks.com/learn/certification/machine-learning-professional>

## Duration

1 days / 7 hours

## Examination description

The training does not conclude with an exam. However, it prepares participants for the Databricks Certified Machine Learning Professional exam.

Exam description: <https://www.databricks.com/learn/certification/machine-learning-professional>