



DP-3014: Build machine learning solutions using Azure Databricks

Microsoft - Data & AI

- **Nível:** Intermédio
 - **Duração:** 7h
-

Sobre o curso

O Azure Databricks é uma plataforma escalável na cloud para análise de dados e machine learning.

Cientistas de dados e engenheiros de machine learning podem utilizar o Azure Databricks para implementar soluções de machine learning em larga escala.

Pré-requisitos

Este curso pressupõe que já tem experiência na utilização de Python para explorar dados e treinar modelos de machine learning com frameworks open source comuns, como Scikit-Learn, PyTorch e TensorFlow.

Programa

- Explorar o Azure Databricks
- Utilizar Apache Spark no Azure Databricks
- Treinar um modelo de machine learning no Azure Databricks
- Utilizar o MLflow no Azure Databricks
- Ajustar hiperparâmetros no Azure Databricks
- Utilizar o AutoML no Azure Databricks
- Treinar modelos de deep learning no Azure Databricks
- Gerir machine learning em produção com o Azure Databricks

Explorar o Azure Databricks

O Azure Databricks é um serviço na cloud que disponibiliza uma plataforma escalável para análise de dados com o Apache Spark.

- Introdução ao Azure Databricks
- Identificar workloads do Azure Databricks
- Compreender os conceitos principais
- Governança de dados com o Unity Catalog e o Microsoft Purview

Utilizar Apache Spark no Azure Databricks

O Azure Databricks baseia-se no Apache Spark e permite que engenheiros e analistas de dados executem tarefas Spark para transformar, analisar e visualizar dados em larga escala.

- Conhecer o Spark
- Criar um cluster Spark
- Utilizar o Spark em notebooks
- Utilizar o Spark para trabalhar com ficheiros de dados
- Visualizar dados

Treinar um modelo de machine learning no Azure Databricks

O machine learning envolve a utilização de dados para treinar um modelo preditivo. O Azure Databricks suporta vários frameworks de machine learning amplamente utilizados que podem ser usados para treinar modelos.

- Compreender os princípios do machine learning
- Machine learning no Azure Databricks
- Preparar dados para machine learning
- Treinar um modelo de machine learning
- Avaliar um modelo de machine learning

Utilizar o MLflow no Azure Databricks

O MLflow é uma plataforma open source para gestão do ciclo de vida de machine learning, com suporte nativo no Azure Databricks.

- Capacidades do MLflow
- Executar experiências com MLflow
- Registrar e disponibilizar modelos com MLflow

Ajustar hiperparâmetros no Azure Databricks

A afinação de hiperparâmetros é uma parte essencial do machine learning. No Azure Databricks, é possível utilizar a biblioteca Hyperopt para otimizar automaticamente os hiperparâmetros.

- Otimizar hiperparâmetros com o Hyperopt
- Rever tentativas do Hyperopt
- Escalar tentativas do Hyperopt

Utilizar o AutoML no Azure Databricks

O AutoML no Azure Databricks simplifica o processo de criação de um modelo eficaz de machine learning para os seus dados.

- O que é o AutoML?
- Utilizar o AutoML na interface do utilizador do Azure Databricks
- Utilizar código para executar uma experiência AutoML

Treinar modelos de deep learning no Azure Databricks

O deep learning utiliza redes neuronais para treinar modelos altamente eficazes para previsões complexas, visão por computador, processamento de linguagem natural e outros workloads de IA.

- Compreender os conceitos de deep learning

- Treinar modelos com o PyTorch
- Distribuir o treino com PyTorch usando o TorchDistributor

Gerir machine learning em produção com o Azure Databricks

O machine learning permite a tomada de decisões baseada em dados e a automação, mas a implementação de modelos em produção para insights em tempo real é um desafio. O Azure Databricks simplifica este processo ao fornecer uma plataforma unificada para criar, treinar e implementar modelos de machine learning em escala, promovendo a colaboração entre cientistas de dados e engenheiros.

- Automatizar as transformações de dados
- Explorar o desenvolvimento de modelos
- Explorar estratégias de implementação de modelos
- Explorar a gestão de versões e do ciclo de vida de modelos