



DP-3014

Implement a Machine Learning solution with Azure Databricks



Información general.

Azure Databricks es una plataforma con escalado en la nube para el análisis de datos y el aprendizaje automático. Los científicos de datos y los ingenieros de aprendizaje automático pueden usar Azure Databricks para implementar soluciones de aprendizaje automático a gran escala.

Requisitos previos.

En esta ruta de aprendizaje se da por supuesto que tiene experiencia en el uso de Python para explorar datos y entrenar modelos de Machine Learning con marcos de código abierto comunes, como Scikit-Learn, PyTorch y TensorFlow.

Duración.

1 Día.

Examen.

Este curso no tiene un examen asociado.

Temario.

Módulo 1: Explorar Azure Databricks.

Azure Databricks es un servicio en la nube que proporciona una plataforma escalable para el análisis de datos mediante Apache Spark.

- Introducción a Azure Databricks.
- Identificación de las cargas de trabajo de Azure Databricks.
- Descripción de los conceptos clave.

Ejercicio: Explorar Azure Databricks.

Módulo 2: Uso de Apache Spark en Azure Databricks.

Azure Databricks se basa en Apache Spark y permite a los ingenieros y analistas de datos ejecutar trabajos de Spark para transformar, analizar y visualizar datos a escala.

- Descubra Spark.
- Creación de un clúster de Spark.
- Uso de Spark en cuadernos.
- Uso de Spark para trabajar con archivos de datos.
- Visualización de datos.

Ejercicio: Uso de Spark en Azure Databricks.

Módulo 3: Entrenamiento de un modelo de Machine Learning en Azure Databricks.

El aprendizaje automático implica el uso de datos para entrenar un modelo predictivo. Azure Databricks admite varios marcos de aprendizaje automático de uso frecuente que puede usar para entrenar modelos.

- Descripción de los principios del aprendizaje automático.
- Aprendizaje automático en Azure Databricks.
- Preparación de los datos para el aprendizaje automático.
- Entrenar un modelo de Machine Learning.
- Evaluar un modelo de Machine Learning.

Ejercicio: Entrenamiento de un modelo de Machine Learning en Azure Databricks.



Módulo 4: Uso de MLflow en Azure Databricks.

MLflow es una plataforma de código abierto para administrar el ciclo de vida de aprendizaje automático que Azure Databricks admite de manera nativa.

- Funcionalidades de MLflow.
- Ejecución de experimentos con MLflow.
- Registro y servicio de modelos con MLflow.

Ejercicio: Uso de MLflow en Azure Databricks.

Módulo 5: Ajuste de hiperparámetros en Azure Databricks.

El ajuste de hiperparámetros es una parte esencial del aprendizaje automático. En Azure Databricks, puede usar la biblioteca Hyperopt para optimizar los hiperparámetros de manera automática.

- Optimización de hiperparámetros con Hyperopt.
- Revisión de las pruebas de Hyperopt.
- Escala de pruebas de Hyperopt.

Ejercicio: Optimización de hiperparámetros para el aprendizaje automático en Azure Databricks.

Módulo 6: Uso de AutoML en Azure Databricks.

AutoML en Azure Databricks simplifica el proceso de creación de un modelo de aprendizaje automático eficaz para sus datos.

- ¿Qué es AutoML?
- Uso de AutoML en la interfaz de usuario de Azure Databricks.
- Uso de código para ejecutar un experimento de AutoML.

Ejercicio: Uso de AutoML en Azure Databricks.

Módulo 7: Entrenamiento de modelos de aprendizaje profundo en Azure Databricks.

El aprendizaje profundo usa redes neuronales para entrenar modelos de aprendizaje automático altamente eficaces para predicciones complejas, Computer Vision, el procesamiento del lenguaje natural y otras cargas de trabajo de IA.

- Comprensión de los conceptos del aprendizaje profundo.
- Entrenamiento de modelos con PyTorch.
- Distribución del entrenamiento de PyTorch con Horovod.

Ejercicio: Entrenamiento de modelos de aprendizaje profundo en Azure Databricks.

