

Data Science Bootcamp

Açıklama

Bu beş gün süren atölye çalışması, Apache Spark 2 ve Hadoop ekosisteminin diğer önemli bileşenlerinin büyük ölçeklerde kullanıldığı veri bilimi ve makine öğrenimi iş akışlarını içerir. Atölye çalışması, gerçek dünyadaki iş zorluklarının üstesinden gelinebilmesi için veri bilimi ve makine öğrenimi yöntemlerinin kullanımının önemini gösterir.

Katılımcılar hayali bir teknoloji şirketinin ve veri setlerinin olduğu senaryoları kullanarak, kritik iş kararlarını desteklemek ve iş yapış şekillerini dönüştüren veri ürünleri geliştirmek üzere gerekli bilgileri keşfetmeye çalışırlar. Materyaller, bir dizi kısa sunumlar, etkileşimli gösteriler, kapsamlı uygulamalı alıştırmalar ve tartışmalar aracılığıyla sunulur.

Atölye çalışması, şu anda Python'u tek makinede küçük veri kümelerinde kullanan ve analizlerini ve makine öğrenimi modellerini dağıtık kümelerdeki büyük veri kümelerine genişletmek ihtiyacı duyan veri bilimcileri için tasarlanmıştır. Veri bilimi ve makine öğrenimi bilgisine sahip veri mühendisleri ve geliştiriciler de bu atölye çalışmasını faydalı bulabilir.

Ön koşullar

Atölye çalışmasına katılanların, Python ile ilgili temel bilgilere sahip olmaları, veri araştırma, veri analizi, istatistiksel veya makine öğrenimi modeli geliştirme gibi konularda deneyimli olmaları gerekir. Hadoop veya Spark ile ilgili bilgi sahibi olunması gerekli değildir.

Eğitim İçeriği

Overview of data science and machine learning at scale

Overview of the Hadoop ecosystem

Working with HDFS data and Hive tables using Hue

Introduction to Cloudera Data Science Workbench

Data Science Bootcamp



Overview of Apache Spark

Reading and writing data

Inspecting data quality

Cleansing and transforming data

Summarizing and grouping data

Combining, splitting, and reshaping data

Exploring data

Configuring, monitoring, and troubleshooting Spark applications

Overview of machine learning in Spark MLlib

Extracting, transforming, and selecting features

Building and evaluating regression models

Building and evaluating classification models

Building and evaluating clustering models

Cross-validating models and tuning hyperparameters

Building machine learning pipelines

Deploying machine learning models