page2image568

Big Data Hadoop and Spark Developer

Lesson-End Project Solution



**Linear Regression with Real-world Dataset**

**Steps to Perform:**

**Step 1:** Log in to your LMS account

**Step 2:** Open the course “**Big Data Hadoop and Spark Developer**”

**Step 3:** Download the datasets from the “**Course Resources**” section

**Step 4:** On the left side, click on the “**PRACTICE LABS**” tab and click on the “**LAUNCH LAB**” button

Graphical user interface, application, Teams

Description automatically generated

**Step 5:** Again, click on the “**LAUNCH LAB**” button

Graphical user interface, application

Description automatically generated

**Step 6:** Click on “HUE” to upload the datasets

Graphical user interface, application, website

Description automatically generated

**Step 7**: Log in to the **HUE** and create a directory named **“data\_files”** and upload the **“customers\_ml”** folder into it

Graphical user interface, text, application, email

Description automatically generated

**Step 8:** Click on **“the Webconsole**” and click on the “**Auth Url**”.

Graphical user interface, application, website

Description automatically generated

**Step 9:** Copy the “**Username**” and the “**Password**” provided to log in to the Web console

**Step 10:** Paste the “**Username**” and the “**Password**” on the console and click on Enter

**Note:** The password will not be visible when pasted on the console.

Text

Description automatically generated

**Step 11:** Log in to the PySpark shell

**Command:**

pyspark3

Text

Description automatically generated

**Step 12:** Import the required packages as shown below:

Text

Description automatically generated

**Step 13:** Read the folder from the HDFS and display the 10 records of the dataset using the below command:

A picture containing graphical user interface

Description automatically generated

**Step 14:** Divide the input features and the label of the dataset to perform the Linear Regression

Graphical user interface, text

Description automatically generated

**Step 15:**  Initialize the Linear Regression model using the below command:

Text

Description automatically generated

**Step 16:** Print the coefficients and intercept of the model with residuals

Text

Description automatically generated

**Step 17:** Print the root mean squared value, mean squared value, and r2 value

Text

Description automatically generated