02:00:00 Start Lab

Feature Analysis Using TensorFlow Data Validation and Facets

** * Rate Lab

Overview Setup Launch Al Platform Notebooks Clone course repo within your Al Platform Notebooks Feature Analysis Using TensorFlow Data Validation and Facets End your lab

user. Errors that result from this bias can disproportionately impact some users more than others.

Overview

2 hours

TensorFlow Data Validation (TFDV) is one tool you can use to analyze your data to find potential problems in your data, such as missing values and data imbalances - that can lead to Fairness disparities. The TFDV tool analyzes training and serving data to compute descriptive statistics, infer a schema, and detect data anomalies. Facets Overview provides a succinct visualization of these statistics for easy browsing. Both the TFDV and Facets are tools that are

Bias can manifest in any part of a typical machine learning pipeline, from an unrepresentative

dataset, to learned model representations, to the way in which the results are presented to the

part of the Fairness Indicators. In this lab, we use TFDV to compute descriptive statistics that provide a quick overview of the data in terms of the features that are present and the shapes of their value distributions. We use Facets Overview to visualize these statistics using the Civil Comments dataset.

• Use TFRecords to load record-oriented binary format data • Use TFDV to generate statistics and Facets to visualize the data

Learning Objectives

· Analyze label distribution for subset groups

Use the TFDV widget to answer questions

- Setup

time block.

Username

cost.

There is no pause feature. You can restart if needed, but you have to start at the beginning.

2. Note the lab's access time (for example, 02:00:00 and make sure you can finish in that

For each lab, you get a new Google Cloud project and set of resources for a fixed time at no

3. When ready, click START LAB

Caution: When you are in the console, do not deviate

from the lab instructions. Doing so may cause your

account to be blocked. Learn more.

1. Make sure you signed into Qwiklabs using an incognito window.

4. Note your lab credentials. You will use them to sign in to the Google Cloud Console. **Open Google Console**

google2876526_student@qwiklabs.n 📋 Password TG959yrKDX GCP Project ID Ê qwiklabs-gcp-0855e773352d3560 New to labs? View our introductory video! 5. Click Open Google Console.

6. Click **Use another account** and copy/paste credentials for **this** lab into the prompts.

If you use other credentials, you'll get errors or incur charges.

7. Accept the terms and skip the recovery resource page.

your work and removes the project.

To launch AI Platform Notebooks:

ARTIFICIAL INTELLIGENCE

Data Labeling

Al Platform

Tables

Step 1

Click on the Navigation Menu. Navigate to AI Platforms, then to Notebooks.

Dashboard

Jobs

Models

On the Notebook instances page, click + NEW INSTANCE . Select TensorFlow 2.x without GPUs.

Notebook instances BETA → ■ NEW INSTANCE C REFRESH ► START ■ STOP 🖰 RESET 👚 DELETE

TensorFlow 1.14 pre-installed with support for Keras

TensorFlow 2.0 pre-installed with support for Keras

0

Without GPUs

With 1 NVIDIA Tesla K80

R 3.6 and key libraries pre-installed

TensorFlow 1.14

TensorFlow 2.0

Do not click End Lab unless you are finished with the lab or want to restart it. This clears

Natural Language Al Hub Notebooks

Launch Al Platform Notebooks

Talent Solution Translation

■ Google Cloud Platform • qwiklabs-gcp-gcpd-2f6ffe6a37af ▼

Vision

Al Platform

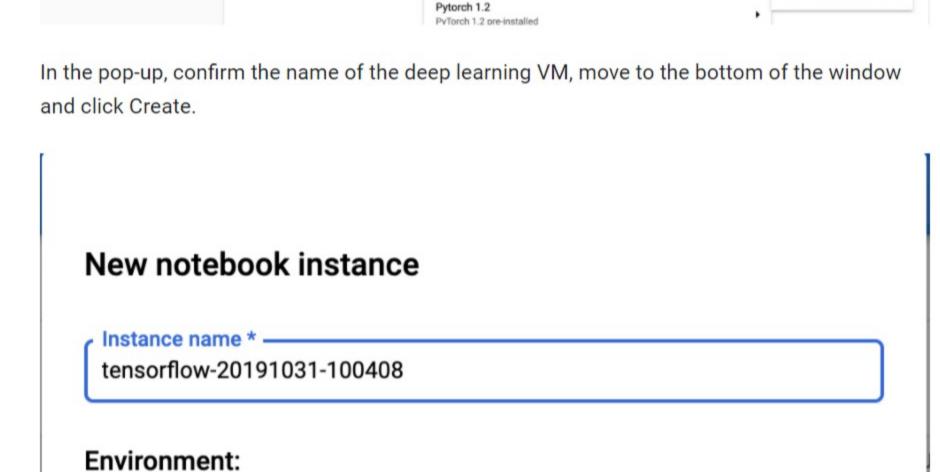
Data Labeling

Step 2

Create and use Jupyter Notebooks with a no JupyterLab pre-installed and are configured

frameworks. Learn more

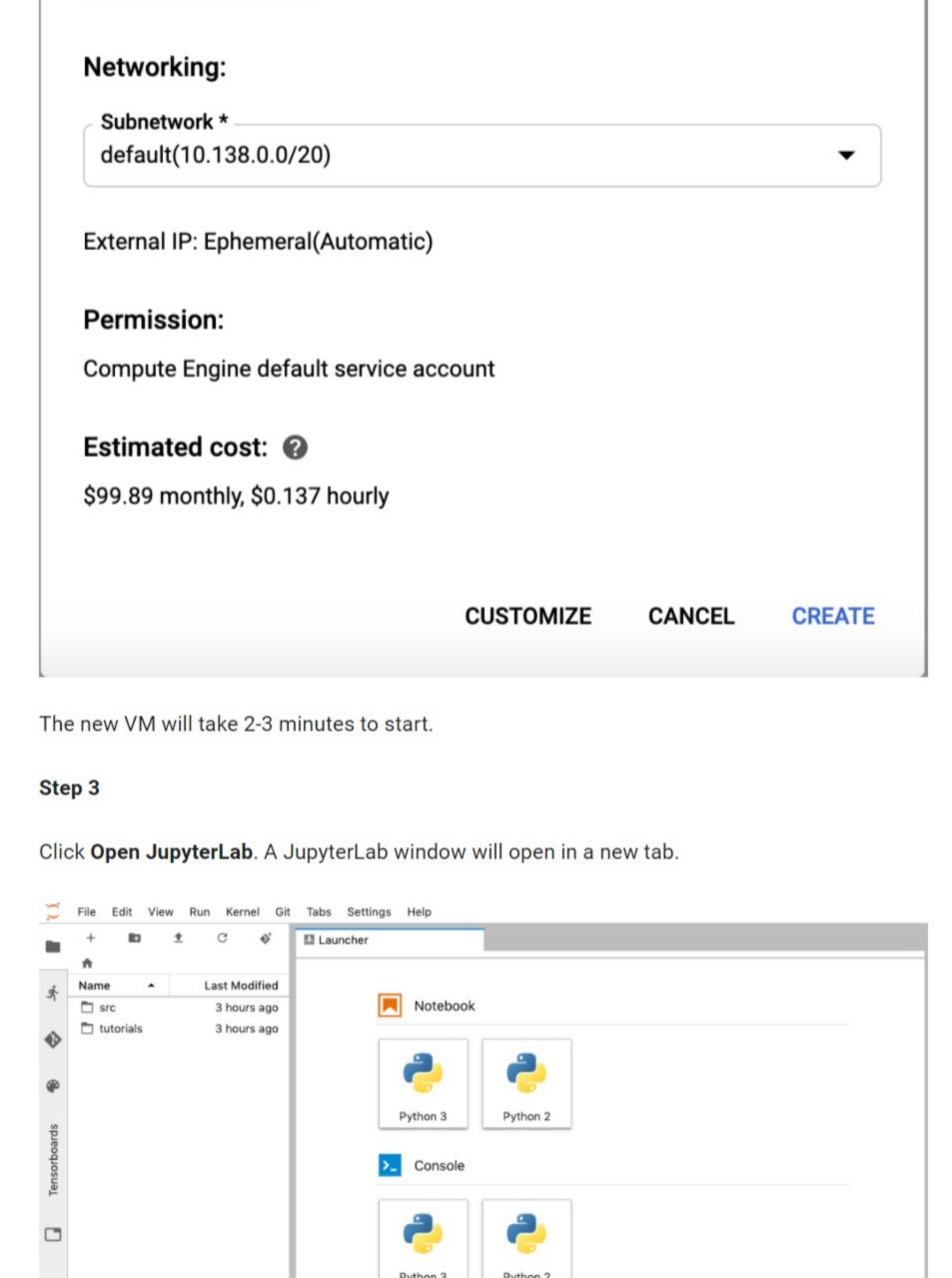
No notebook instances to display



Region and zone: us-west1-b Machine type: 4 vCPUs, 15 GB RAM Boot disk: 100 GB Disk

Image: TensorFlow 2.0 (with Intel® MKL-DNN/MKL and CUDA 10.0)

Packages: python2, python3, scikit-learn, pandas, and nltk.



Other

\$_

Clone course repo within your Al Platform

To clone the training-data-analyst notebook in your JupyterLab instance:

Text File

Tensorboard

Step 1 In JupyterLab, click the Terminal icon to open a new terminal.

Notebook

Python 3

Python 3

Console

Launcher

Notebooks instance

Python 2

Python 2

Terminal Text File Tensorboard Step 2 At the command-line prompt, type in the following command and press Enter. git clone https://github.com/GoogleCloudPlatform/training-data-analyst Step 3 Confirm that you have cloned the repository by double clicking on the training-dataanalyst directory and ensuring that you can see its contents. The files for all the Jupyter notebook-based labs throughout this course are available in this directory. Edit View Run Kernel Git C **Last Modified** Name blogs 2 minutes ago bootcamps 2 minutes ago courses 2 minutes ago □ CPB100 2 minutes ago datalab 2 minutes ago

Feature Analysis Using TensorFlow Data

quests

self-paced-l...

CONTRIBU...

README.md

Validation and Facets

LICENSE

2 minutes ago

Step 1 In the notebook interface, navigate to training-data-analyst > courses > machine_learning > deepdive2 > introduction_to_tensorflow > labs and open adv_tfdv_facets.ipynb.

In the notebook interface, click on Edit > Clear All Outputs (click on Edit, then in the drop-down

menu, select Clear All Outputs). Carefully read through the notebook instructions and fill in lines marked with #TODO where you need to complete the code as needed

Step 2

found in the notebook UI under Run. • Hints may also be provided for the tasks to guide you along. Highlight the text to read the hints (they are in white text).

• If you need more help, you may take a look at the complete solution by navigating to

training-data-analyst > courses > machine_learning > deepdive2 >

introduction_to_tensorflow > solutions and open adv_tfdv_facets.ipynb.

Tip: To run the current cell you can click the cell and hit shift+enter. Other cell commands are

- End your lab When you have completed your lab, click End Lab. Qwiklabs removes the resources you've
- You will be given an opportunity to rate the lab experience. Select the applicable number of stars, type a comment, and then click Submit. The number of stars indicates the following:
- 2 stars = Dissatisfied • 3 stars = Neutral • 4 stars = Satisfied • 5 stars = Very satisfied

used and cleans the account for you.

• 1 star = Very dissatisfied

- You can close the dialog box if you don't want to provide feedback. For feedback, suggestions, or corrections, please use the Support tab.
- ©2020 Google LLC All rights reserved. Google and the Google logo are trademarks of Google LLC. All other company and product names may be trademarks of the respective companies with which they are associated.