



ML on Terra

amyu@verily.com
15th Sept., 2022

These slides: bit.ly/ml-terra-sept15





“Physics requires math in the same way that biology will require AI.”

Introduction and logistics



- A little about me
- About this webinar & what it will cover
- Example workspace:
<https://app.terra.bio/#workspaces/verily-terra-solutions/ml-on-terra> & related set of [articles](#) & [GH repo](#)
- Some things to pay attention to when working through the examples yourself later:
 - Clone the example workspace, and work in the clone
 - Use the indicated Cloud Environment config for the in-notebook training examples
 - For many of the examples, a [‘native’ GCP account](#) is accessed. The 00_pcam_setup notebook walks you through the setup process for that.
- How get help later & give feedback on the examples

These slides: bit.ly/ml-terra-sept15



Agenda



- Introduction and logistics
- In-notebook model training
- In-notebook training with hyperparameter tuning
- Training and serving models via [Vertex AI](#);
+ Experiments API & Managed TensorBoard
- [BigQuery ML](#) (BQML): Automatically train models from BQ datasets
- [time permitting] [Vertex AI Pipelines](#)
- Discussion

These slides: bit.ly/ml-terra-sept15

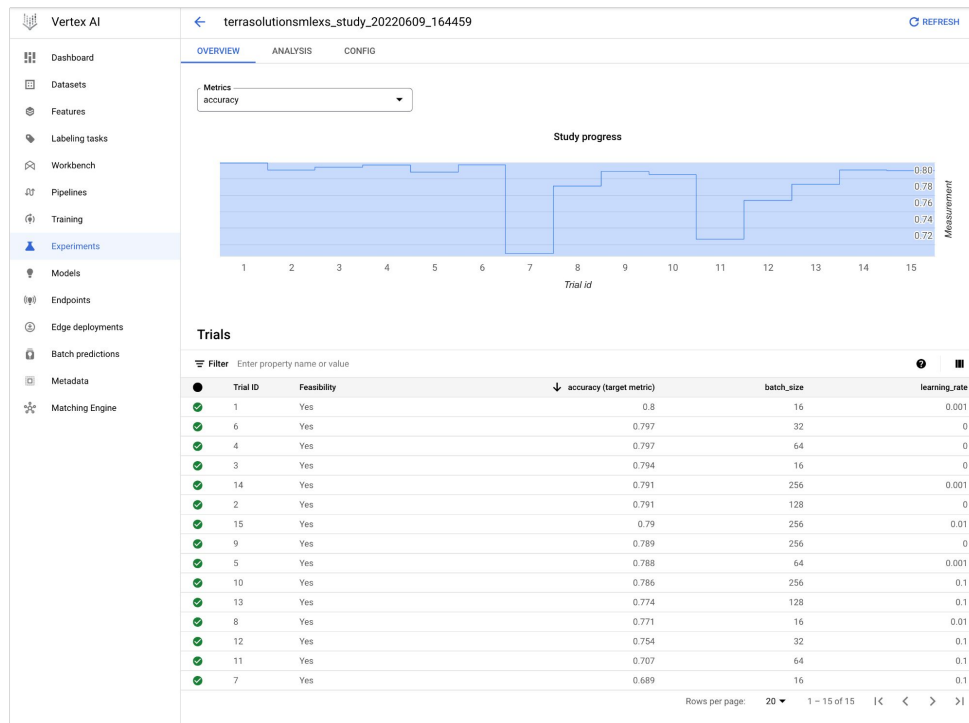


In-notebook model training

- Workspace notebook: 01_keras_pcam.ipynb
- Article:
<https://support.terra.bio/hc/en-us/articles/5496810552987-Machine-learning-in-Terra-Part-I-Train-in-a-Terra-notebook>

In-notebook training with HP tuning

- [Workspace notebook:](#)
04_vizier_pcam_model_training.ipynb



Vizier: “study” trials

Trial : batch_size is 16.0, learning_rate is 0.1.

===== Start Trial: [projects/518704281671/locations/us-central1/studies/3562191698396/trials/1] =====

using batch_size 16; learning_rate 0.1

gs://fc-5bb10fd8-a839-44b6-a6b7-4006b6216a26/logs/pcam/terasolutionsmlexs_study_20220907_160704/1

20220907160759

INFO:tensorflow:Using MirroredStrategy with devices ('/job:localhost/replica:0/task:0/device:GPU:0',)

INFO:tensorflow:Using MirroredStrategy with devices ('/job:localhost/replica:0/task:0/device:GPU:0',)

Number of devices: 1

Model: "model_1"

Layer (type)	Output Shape	Param #
=====		
input_4 (InputLayer)	[(None, 96, 96, 3)]	0
rescaling_1 (Rescaling)	(None, 96, 96, 3)	0
xception (Functional)	(None, 3, 3, 2048)	20861480
global_average_pooling2d_1 (GlobalAveragePooling2D)	(None, 2048)	0
dropout_1 (Dropout)	(None, 2048)	0
dense_1 (Dense)	(None, 2)	4098

=====

Total params: 20,865,578

Trainable params: 4,098

Non-trainable params: 20,861,480

gs://fc-5bb10fd8-a839-44b6-a6b7-4006b6216a26/logs/pcam/terasolutionsmlexs_study_20220907_160704/1

./checkpoints/20220907160759/checkpoints

layer_config = serialize_layer_fn(layer)

Epoch 1/2

2022-09-07 16:08:08.094919: I tensorflow/stream_executor/cuda/cuda_dnn.cc:366] Loaded cuDNN version 8200

16384/16384 [=====] - ETA: 0s - loss: 8.7012 - accuracy: 0.7476

...

Trial : batch_size is 32.0, learning_rate is 0.01.

===== Start Trial: [projects/518704281671/locations/us-central1/studies/3562191698396/trials/2] =====

using batch_size 32; learning_rate 0.01

gs://fc-5bb10fd8-a839-44b6-a6b7-4006b6216a26/logs/pcam/terasolutionsmlexs_study_20220907_160704/2

20220907162046

INFO:tensorflow:Using MirroredStrategy with devices ('/job:localhost/replica:0/task:0/device:GPU:0',)

INFO:tensorflow:Using MirroredStrategy with devices ('/job:localhost/replica:0/task:0/device:GPU:0',)

Number of devices: 1

Model: "model_3"

Layer (type)	Output Shape	Param #
=====		
input_8 (InputLayer)	[(None, 96, 96, 3)]	0
rescaling_3 (Rescaling)	(None, 96, 96, 3)	0
xception (Functional)	(None, 3, 3, 2048)	20861480
global_average_pooling2d_3 (GlobalAveragePooling2D)	(None, 2048)	0
dropout_3 (Dropout)	(None, 2048)	0
dense_3 (Dense)	(None, 2)	4098

=====

Total params: 20,865,578

Trainable params: 4,098

Non-trainable params: 20,861,480

gs://fc-5bb10fd8-a839-44b6-a6b7-4006b6216a26/logs/pcam/terasolutionsmlexs_study_20220907_160704/2

./checkpoints/20220907162046/checkpoints

Epoch 1/2

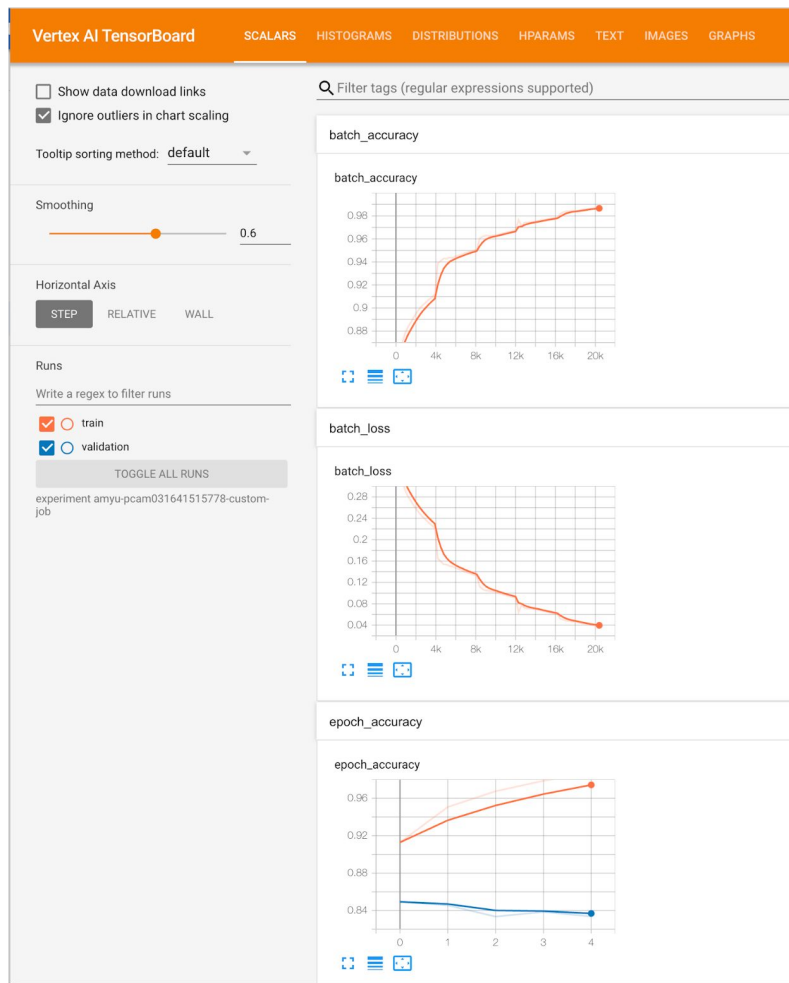
8192/8192 [=====] - ETA: 0s - loss: 0.7626 - accuracy: 0.7700INFO

...

Training and serving with Vertex AI: + Experiments API, managed TensorBoard

- [Workspace notebook: 02_1_vertex_ai_pcam.ipynb](#)
- Article:
<https://support.terra.bio/hc/en-us/articles/5510714723611-Machine-learning-on-Terra-Part-II-Scaling-training-and-serving>
- Additional related materials are in [this notebook](#).
- Vertex AI docs:
 - [Training AutoML Models | Vertex AI | Google Cloud](#)
 - [Introduction to Vertex AI Experiments | Google Cloud](#)
 - [Request predictions | Vertex AI | Google Cloud](#)

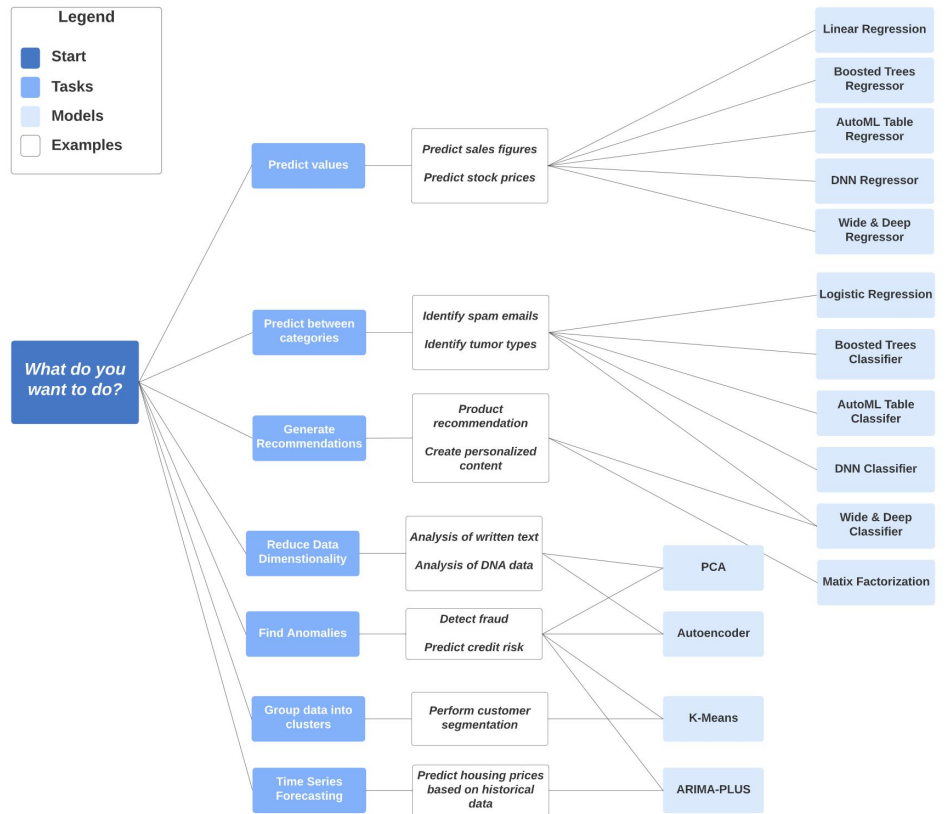
Managed TensorBoard



BigQuery ML (BQML)

- [Notebook \(bit.ly/3d5lQ4S\)](https://bit.ly/3d5lQ4S)
- BQML documentation: <https://cloud.google.com/bigquery-ml/docs>
- [List of supported model types](#)
- More examples: [this tutorial](#) and [others](#)

Model selection guide



BQML console

The screenshot displays the Google Cloud BQML console interface. The top navigation bar includes the Google Cloud logo, a user profile 'amy-jo-demos', and a search bar. The left sidebar shows the 'Explorer' view with a search bar and a list of pinned projects under 'amy-jo-demos', including 'External connections', 'amyu_test', 'aunruh_testds', 'covid19_demos', 'covid_demos', 'mimic_models', and 'Models (7)'. The main panel shows the details for the model 'predict_mortality_diag_128'. The 'DETAILS' tab is active, displaying the model type as 'LOGISTIC_REGRESSION' and the data location as 'US'. Below this, the 'Model details' section provides information such as the Model ID, Description, Labels, Date created, Model expiry, Date modified, Data location, Model type, and Loss type. The 'Training options' section explains that these are optional parameters added to the script and lists various parameters like Max allowed iterations, Actual iterations, L1 regularisation, L2 regularisation, Early stop, Min relative progress, Learn rate strategy, Line search initial learn rate, and Calculate p-values.

Google Cloud amy-jo-demos Search Products, resources, docs (/)

predict_mortality_diag_128

predict_mortality_diag_128

QUERY MODEL DELETE MODEL EXPORT MODEL

DETAILS TRAINING EVALUATION SCHEMA

Model type
LOGISTIC_REGRESSION

Data location
US

Model details EDIT

Model ID
amy-jo-demos.mimic_models.predict_mortality_diag_128

Description

Labels

Date created
Wednesday, 7 September 2022 at 14:44:18 GMT-07:00

Model expiry
Never

Date modified
Wednesday, 7 September 2022 at 14:44:18 GMT-07:00

Data location
US

Model type
LOGISTIC_REGRESSION

Loss type
Mean log loss

Training options

Training options are the optional parameters that were added in the script to create this model.

Max allowed iterations
20

Actual iterations
11

L1 regularisation
2.00

L2 regularisation
0.00

Early stop
true

Min relative progress
0.01

Learn rate strategy
Line search

Line search initial learn rate
0.10

Calculate p-values
false

Vertex AI Pipelines

- [Workspace notebook](#)
- Article:
<https://support.terra.bio/hc/en-us/articles/5498395787291-Machine-learning-on-Terra-Part-III-ML-Ops>
- Vertex AI documentation:
 - [Vertex AI Pipelines | Google Cloud](#)

Vertex AI Pipelines

Vertex AI

Dashboard

Datasets

Features

Labeling tasks

Workbench

Pipelines

Training

Experiments

Models

Endpoints

Edge deployments

Batch predictions

Metadata

← keras-patchcamelyon-pipeline-20211221174603

CLONE STOP DELETE

① Runtime Graph ✓ 6/6 steps completed Expand Artifacts 100%

Pipeline run analysis

SUMMARY **NODE INFO**

Artifact Info
[VIEW LINEAGE](#)

Name metricsc
Type system.ClassificationMetrics
URI gs://terra-solutions-ml-exs-central1/pipeline_root/aj/518704281671/keras-patchcamelyon-pipeline-20211221174603/classif-model-eval-metrics_-4802706372544167936/metricsc

Confusion matrix Item counts

This table shows how often the model classified each label correctly (in blue), and which labels were most often confused for that label (in gray).

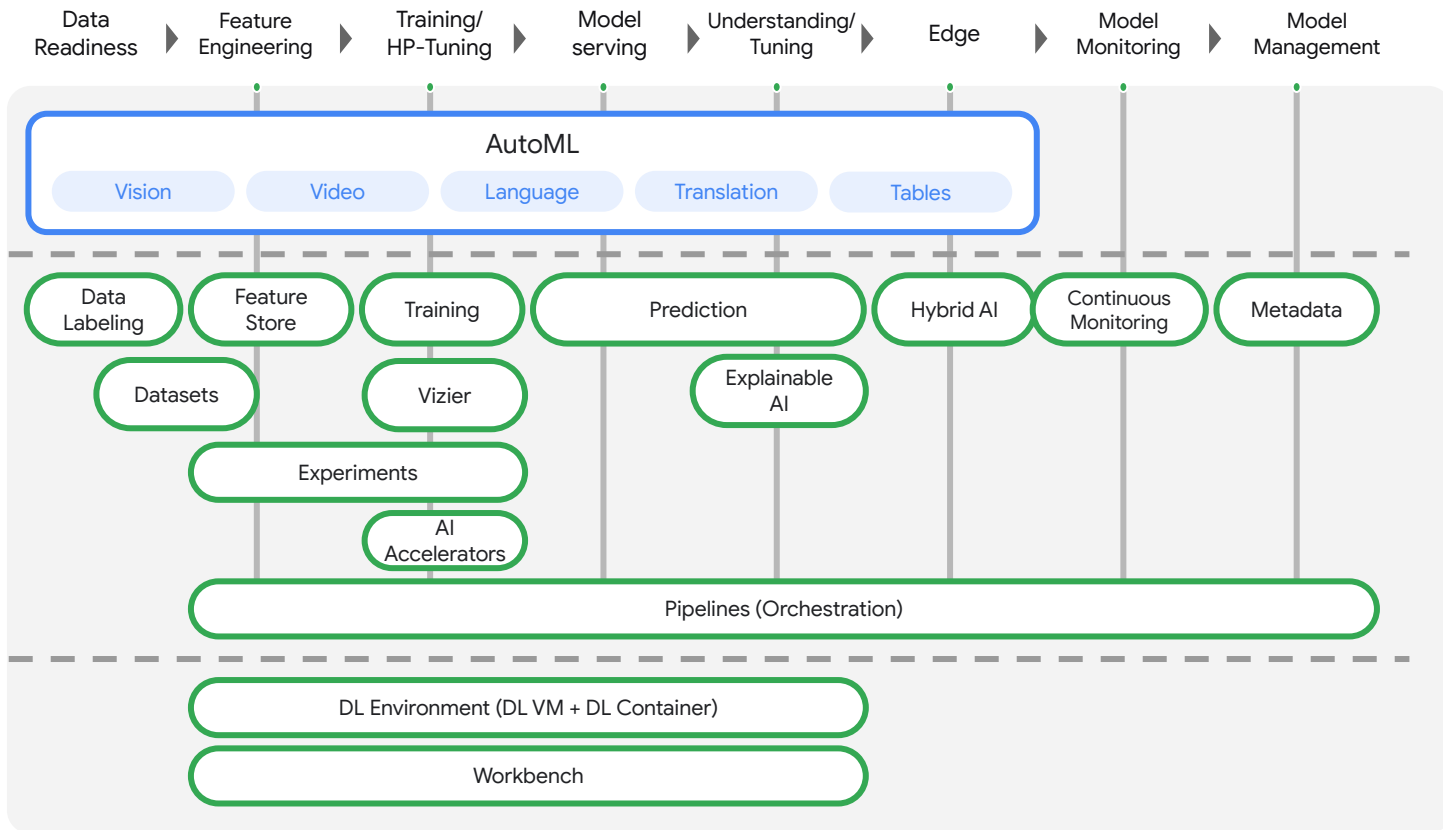
True label	Predicted label non_metastatic	metastatic
non_metastatic	96%	4%
metastatic	28%	72%

More Resources

More resources and relevant services

- These slides: bit.ly/ml-terra-sept15
- [Terra support site](#)
- [Vertex AI documentation](#)
- Building models using [AutoML](#) + your own data (for tabular, text, image, and video datasets)
- [Vertex Explainable AI](#) : helps you understand how data features contributed to a prediction result
- [AlphaFold batch inference with Vertex AI Pipelines](#)
- [PyTorch on GCP](#)
- Cloud [Dataflow's](#) new [RunInference transform](#)

What's included in Vertex AI?



end