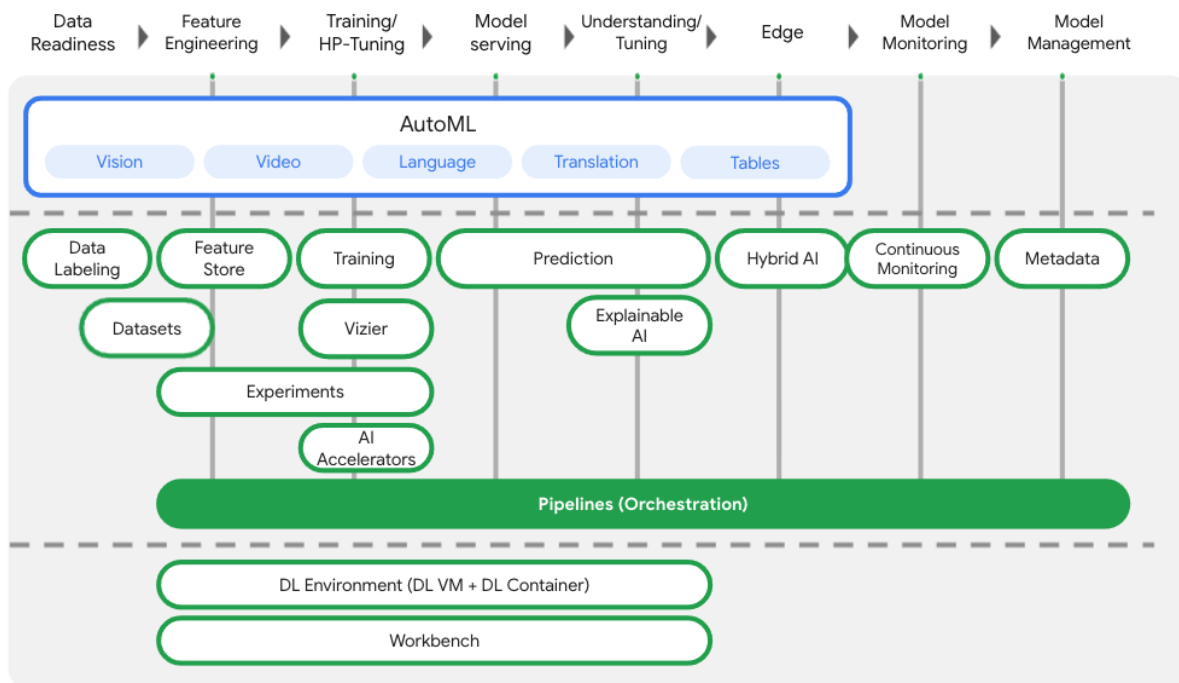


a) Vertex AI pipeline Intro

Reference: <https://codelabs.developers.google.com/vertex-pipelines-intro#0>

Objectives:

- Use the Kubeflow Pipelines SDK to build scalable ML pipelines
- Create and run a 3-step intro pipeline that takes text input
- Create and run a pipeline that trains, evaluates, and deploys an AutoML classification model
- Use pre-built components for interacting with Vertex AI services, provided through the `google_cloud_pipeline_components` library
- Schedule a pipeline job with Cloud Scheduler



Cloud environment setup:

Enable API services and create google bucket:

Vertex AI

hello-world-pipeline-20211125173807

Runtime Graph: 0/3 steps completed

Build-sentence (python 3.7)

Pipeline run analysis

Execution Info: Pending

Parameter	Type	Value
emoji	string	<undefined_till_runtime>
product	string	<undefined_till_runtime>
emojixtext	string	<undefined_till_runtime>

Creating an end to end ML pipeline

```

import google.cloud.aiplatform as aiplatform
import json
import os

project = "my-project"
location = "us-central1"
display_name = "train-automl-beans"

gcc_aip.ModelDeployOp(
    model=training_op.outputs["model"],
    endpoint=endpoint_op.outputs["endpoint"],
    dedicated_resources_min_replica_count=1,
    dedicated_resources_max_replica_count=1,
    dedicated_resources_machine_type="n1-standard-4",
)

[20]: compiler.Compiler().compile(
    pipeline_func=pipeline, package_path="tab_classif_pipeline.json"
)

Defining the job

[21]: ml_pipeline_job = aiplatform.PipelineJob(
    display_name="automl-tab-beans-training",
    template_path="tab_classif_pipeline.json",
    pipeline_root=PIPELINE_ROOT,
    parameter_values={"project": PROJECT_ID, "display_name": DISPLAY_NAME},
    enable_caching=True
)

[22]: ml_pipeline_job.submit()

INFO:google.cloud.aiplatform.pipeline_jobs:Creating PipelineJob
INFO:google.cloud.aiplatform.pipeline_jobs:PipelineJob created. Resource name: projects/714833773523/locations/us-central1/pipelineJobs/automl-tab-beans-training-v2-28211125174651
INFO:google.cloud.aiplatform.pipeline_jobs:To use this PipelineJob in another session:
INFO:google.cloud.aiplatform.pipeline_jobs:pipeline_job = aiplatform.PipelineJob.get('projects/714833773523/locations/us-central1/pipelineJobs/automl-tab-beans-training-v2-28211125174651')
INFO:google.cloud.aiplatform.pipeline_jobs:View Pipeline Job:
https://console.cloud.google.com/vertex-ai/locations/us-central1/pipelines/runs/automl-tab-beans-training-v2-28211125174651?project=714833773523

```

Compiling and running the pipeline

console.cloud.google.com/vertex-ai/locations/us-central1/pipelines/runs/automl-tab-beans-training-v2-20211125174651?authuser=3&project=cmpe260

Google Cloud Platform cmpe260 Search products and resources

Vertex AI

- Dashboard
- Datasets
- Features
- Labeling tasks
- Workbench
- Pipelines
- Training
- Experiments
- Models
- Endpoints
- Batch predictions
- Metadata

Marketplace

automl-tab-beans-training-v2-20211125174651

Runtime Graph 6/6 steps completed Expand Artifacts 95%

tabulardataset-create
google.VertexDataset

automltabulartrainingjob-
google.VertexModel

classification-model-eval-
google.VertexModel

condition-deploy-decision-
google.VertexModel

endpoint-create
google.VertexEndpoint

Pipeline run analysis

SUMMARY NODE INFO

Artifact Info

VIEW LINEAGE

Name endpoint
Type google.VertexEndpoint
URI <https://us-central1-aiplatform.googleapis.com/v1/projects/cmpe260/locations/us-central1/endpoints/3070760054517923840>

Properties

All properties of the artifact.

resourceName projects/cmpe260/locations/us-central1/endpoints/3070760054517923840

Logs

Data Set and End Point View

console.cloud.google.com/vertex-ai/locations/us-central1/metadata-stores/default/artifacts/1676755780854661555?authuser=3&project=cmpe260

Google Cloud Platform cmpe260 Search products and resources

Vertex AI

- Dashboard
- Datasets
- Features
- Labeling tasks
- Workbench
- Pipelines
- Training
- Experiments
- Models
- Endpoints
- Batch predictions
- Metadata

Marketplace

endpoint

Lineage Graph Expand Executions 100%

dataset
google.VertexDataset

model
google.VertexModel

endpoint
google.VertexEndpoint

metrics
system.Metrics

metricsc
system.ClassificationMetrics

Confusion matrix

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	DERMASON	SIRA	SEKER	HOROZ
DERMASON	93%	5%	2%	—
SIRA	7%	91%	1%	2%
SEKER	1%	3%	96%	—
HOROZ	1%	3%	—	96%
CALI	—	1%	—	4%
BARBUNYA	—	4%	1%	—
BOMBAY	—	—	—	—

console.cloud.google.com/vertex-ai/endpoints?authuser=3&project=cme260

Google Cloud Platform

Vertex AI

Endpoints

CREATE ENDPOINT

Refresh

Endpoints are machine learning models made available for online prediction requests. Endpoints are useful for timely predictions from many users (for example, in response to an application request). You can also request batch predictions if you don't need immediate results.

To create an endpoint, you need at least one machine learning model. [Learn more](#)

Region: us-central1 (Iowa)

1 selected EDIT LABELS DELETE

Name	ID	Status	Models	Region	Monitoring	Most recent alerts	Last updated	API	Notification	Labels
train-automi-beans	3070780054517923840	Active	1	us-central1	Disabled	—	Nov 25, 2021, 12:19:55 PM	Sample request		

Model Output

console.cloud.google.com/vertex-ai/locations/us-central1/models/3476009254190579712/evaluate?authuser=3&project=cme260

Google Cloud Platform

Vertex AI

automi-beans1637862305

EVALUATE DEPLOY & TEST

VIEW DATASET EXPORT

Filter labels

Confidence threshold: 0.5

All labels

Label	Count
BOMBAY	1
SEKER	0.98905
DERMASON	0.98528
CALI	0.97961
BARBUNYA	0.97727
HORROZ	0.97729
SIRA	0.94841

PR AUC: 0.98
ROC AUC: 0.994
Log loss: 0.2
F1 score: 0.9304574
Precision: 92.6%
Recall: 92.6%
Created: Nov 25, 2021, 12:13:21 PM

To evaluate your model, set the confidence threshold to see how precision and recall are affected. The best confidence threshold depends on your use case. Read more [about metrics](#) to learn how evaluation metrics can be used.

Precision-recall curve

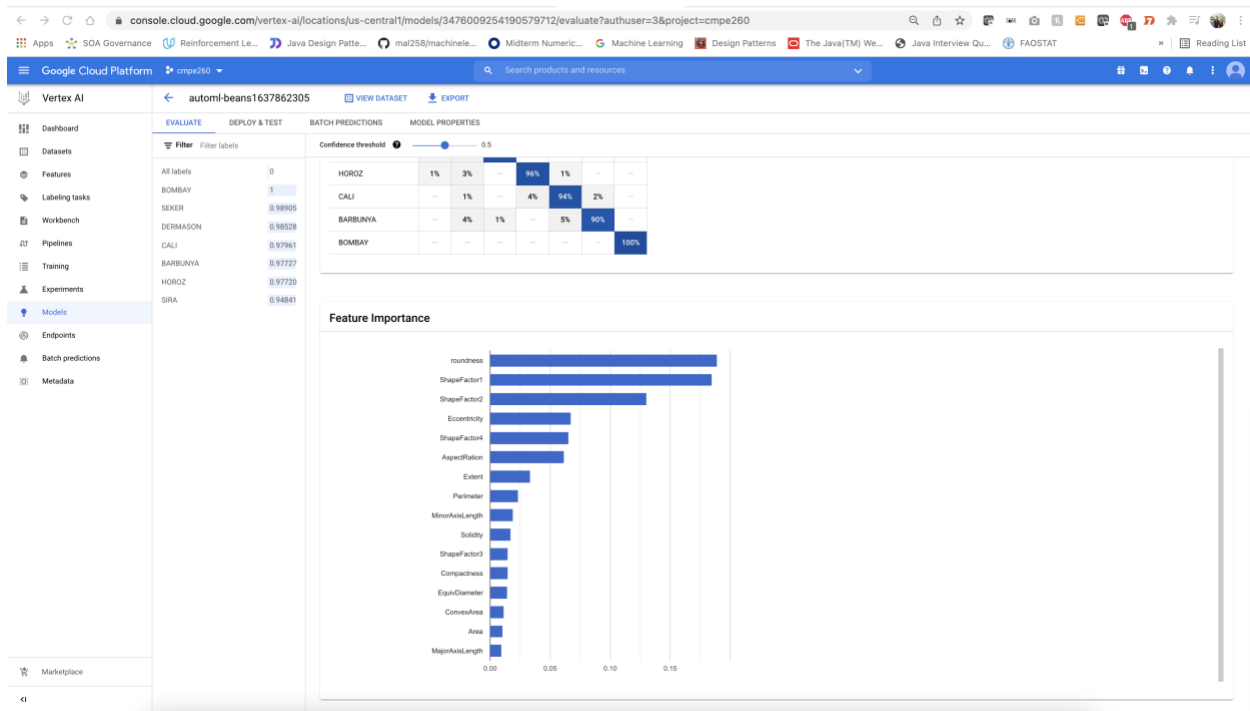
ROC curve

Precision-recall by threshold

Confusion matrix

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	DERMASON	SIRA	SEKER	HORROZ	CALI	BARBUNYA	BOMBAY
DERMASON		93%	5%	2%	—	—	—	—
SIRA		7%	91%	1%	2%	—	0%	—
SEKER		1%	3%	96%	—	—	—	—
HORROZ		1%	3%	—	96%	1%	—	—
CALI		—	1%	—	4%	94%	2%	—
BARBUNYA		—	4%	1%	—	5%	90%	—
BOMBAY		—	—	—	—	—	—	100%



1f5f3640670c33b-dot-us-central1.notebooks.googleusercontent.com/lab/tree/First_Pipeline.ipynb

File Edit View Run Kernel Git Tabs Settings Help

Filter files by name

Name	Last Modified
/	
src	3 hours ago
tutorials	3 hours ago
First_Pipeline.ipynb	seconds ago
first-component.yaml	3 hours ago
intro_pipeline_job.json	3 hours ago
tab_classif_pipeline.json	3 hours ago
tabular_eval_component.yaml	3 hours ago

First_Pipeline.ipynb

```
gcc_aip.ModelDeployOp(  
    model=train_op.outputs["model"],  
    endpoint_endpoint=train_op.outputs["endpoint"],  
    dedicated_resources_min_replica_count=1,  
    dedicated_resources_max_replica_count=1,  
    dedicated_resources_machine_type="n1-standard-4",  
)  
  
Compiling and running the end to end pipeline  
  
[20]: compiler.Compiler().compile(  
    pipeline_func=pipeline, package_path="tab_classif_pipeline.json"  
)  
  
Defining the job  
  
[21]: ml_pipeline_job = aiplatform.PipelineJob(  
    display_name="automi-tab-beans-training",  
    template_path="tab_classif_pipeline.json",  
    pipeline_root=PIPELINE_ROOT,  
    parameter_values={"project": PROJECT_ID, "display_name": DISPLAY_NAME},  
    enable_caching=True  
)  
  
[22]: ml_pipeline_job.submit()  
  
INFO:google.cloud.aiplatform.pipeline_jobs:Creating PipelineJob  
INFO:google.cloud.aiplatform.pipeline_jobs:PipelineJob created. Resource name: projects/714833773523/locations/us-central1/pipelineJobs/automi-tab-beans-training-v2-2821125174651  
INFO:google.cloud.aiplatform.pipeline_jobs:To use this PipelineJob in another session:  
INFO:google.cloud.aiplatform.pipeline_jobs:PipelineJob = aiplatform.PipelineJob.get('projects/714833773523/locations/us-central1/pipelineJobs/automi-tab-beans-training-v2-2821125174651')  
INFO:google.cloud.aiplatform.pipeline_jobs:View Pipeline Job:  
https://console.cloud.google.com/vertex-ai/locations/us-central1/pipelines/runs/automi-tab-beans-training-v2-2821125174651?project=714833773523  
  
[23]: pipeline_df = aiplatform.get_pipeline_df(pipeline="automi-tab-beans-training-v2")  
small_pipeline_df = pipeline_df.head(2)  
small_pipeline_df  
  
[23]:
```

pipeline_name	run_name	param.input:display_name	param.input:thresholds_dict_str	param.input:bq_source	param.input:gcp_region	param.input:api_endpoint	param.input:pro
automi-tab-beans-training-v2	automi-tab-beans-training-v2-20211125174651	automi-beans1637862305	("auRoc": 0.95)	bq://aij-dev-demos.beans.beans1	us-central1	aiplatform.googleapis.com	cme

Simple 0 1 No Kernel | Idle

Mode: Edit Ln 1, Col 1 First_Pipeline.ipynb