

# Intro to Vertex Pipelines

This document shows the details of creating and running ML Pipelines

## 1. Setting the project and enabling APIs required to build vertex pipelines

```
Welcome to Cloud Shell! Type "help" to get started.
Your Cloud Platform project in this session is set to cmpe297project.
Use "gcloud config set project [PROJECT_ID]" to change to a different project.
arpithagurumurthy_ml@cloudshell:~ (cmpe297project)$ gcloud auth list
Credentialed Accounts

ACTIVE: *
ACCOUNT: arpithagurumurthy.ml@gmail.com

To set the active account, run:
  $ gcloud config set account `ACCOUNT`

arpithagurumurthy_ml@cloudshell:~ (cmpe297project)$ gcloud config list project
[core]
project = cmpe297project

Your active configuration is: [cloudshell-16932]
arpithagurumurthy_ml@cloudshell:~ (cmpe297project)$ echo $GOOGLE_CLOUD_PROJECT
cmpe297project
arpithagurumurthy_ml@cloudshell:~ (cmpe297project)$ gcloud services enable compute.googleapis.com
> containerregistry.googleapis.com \
> aiplatform.googleapis.com \
> cloudbuild.googleapis.com \
> cloudfunctions.googleapis.com

Operation "operations/acf.p2-267810977261-0805e224-bdc0-48e6-bbff-4351517f6efb" finished successfully.
arpithagurumurthy_ml@cloudshell:~ (cmpe297project)$
```

## 2. Creating a Cloud Storage Bucket

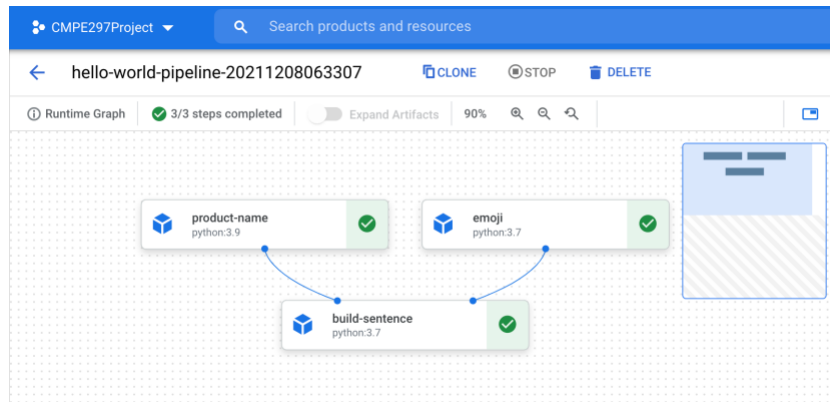
```
arpithagurumurthy_ml@cloudshell:~ (cmpe297project)$ BUCKET_NAME=gs://$GOOGLE_CLOUD_PROJECT-bucket
arpithagurumurthy_ml@cloudshell:~ (cmpe297project)$ gsutil mb -l us-central1 $BUCKET_NAME
Creating gs://cmpe297project-bucket/.
arpithagurumurthy_ml@cloudshell:~ (cmpe297project)$ gcloud projects describe $GOOGLE_CLOUD_PROJECT > project-info.txt
arpithagurumurthy_ml@cloudshell:~ (cmpe297project)$ PROJECT_NUM=$(cat project-info.txt | sed -nre 's:.projectNumber: (.*)\:lp')
arpithagurumurthy_ml@cloudshell:~ (cmpe297project)$ SVC_ACCOUNT="$(PROJECT_NUM/\)/-compute@developer.gserviceaccount.com"
arpithagurumurthy_ml@cloudshell:~ (cmpe297project)$ gcloud projects add-iam-policy-binding $GOOGLE_CLOUD_PROJECT --member serviceAccount:$SVC_ACCOUNT --role roles/storage.objectAdmin
Updated IAM policy for project [cmpe297project].
bindings:
- members:
  - serviceAccount:service-267810977261@gcp-sa-aiplatform.iam.gserviceaccount.com
    role: roles/aiplatform.serviceAgent
- members:
  - serviceAccount:267810977261@cloudbuild.gserviceaccount.com
    role: roles/cloudbuild.builds.builder
- members:
  - serviceAccount:service-267810977261@gcp-sa-cloudbuild.iam.gserviceaccount.com
    role: roles/cloudbuild.serviceAgent
- members:
  - serviceAccount:service-267810977261@gcp-admin-robot.iam.gserviceaccount.com
    role: roles/cloudfunctions.serviceAgent
- members:
  - serviceAccount:service-267810977261@compute-system.iam.gserviceaccount.com
    role: roles/compute.serviceAgent
- members:
  - serviceAccount:service-267810977261@containerregistry.iam.gserviceaccount.com
    role: roles/containerregistry.ServiceAgent
- members:
  - serviceAccount:267810977261-compute@developer.gserviceaccount.com
  - serviceAccount:267810977261@cloudservices.gserviceaccount.com
  - serviceAccount:cmpe297project@appspot.gserviceaccount.com
    role: roles/editor
- members:
  - serviceAccount:cmpe-297-vertexai@cmpe297project.iam.gserviceaccount.com
  - user:arpithagurumurthy.ml@gmail.com
    role: roles/owner
- members:
  - serviceAccount:service-267810977261@gcp-sa-pubsub.iam.gserviceaccount.com
    role: roles/pubsub.serviceAgent
- members:
  - serviceAccount:267810977261-compute@developer.gserviceaccount.com
    role: roles/storage.objectAdmin
etag: BwXSnDcnS0=
version: 1
arpithagurumurthy_ml@cloudshell:~ (cmpe297project)$
```

## 3. Creating a Vertex AI Workbench instance

Filter Enter property name or value										
<input type="checkbox"/>	<input checked="" type="checkbox"/>	Notebook name ↑		Zone	Auto-upgrade	Environment	Machine type	GPUs	Permission	Last mc
<input type="checkbox"/>	<input checked="" type="checkbox"/>	tensorflow-2-3-20211207-221539	<a href="#">OPEN JUPYTERLAB</a>	us-west1-b	—	TensorFlow:2.3	4 vCPUs, 15 GB RAM	None	Service account	—

**\*\*Attaching the Python notebook for the pipeline in github**

## Pipeline after completion



**\*\*After clicking on the build-sentence component to see the final output**

# Pipeline run analysis

SUMMARY

NODE INFO

Execution Info

✓

Completed

VIEW JOB

VIEW LOGS

Display name	build-sentence
Name	build-sentence
Type	system.ContainerExecution
Duration	2 min 43 sec
Started	Dec 7, 2021, 10:37:05 PM
Completed	Dec 7, 2021, 10:39:48 PM

Input Parameters

Parameter	Type	Value
product	string	Vertex Pipelines
emojitext	string	sparkles
emoji	string	🔥

Output Parameters

Parameter	Type	Value
Output	string	Vertex Pipelines is 🔥

### Creating an end-to-end ML pipeline

- Using UCI Machine Learning [Dry beans dataset](#), from: KOKLU, M. and OZKAN, I.A., (2020), "Multiclass Classification of Dry Beans Using Computer Vision and Machine Learning Techniques."

This pipeline will:

- Create a dataset in Vertex AI
- Train a tabular classification model with AutoML
- Get evaluation metrics on this model
- Deploy the model to an endpoint

**\*\*Attaching the Python notebook for the pipeline in github**

Output:

INFO:google.cloud.aiplatform.pipeline\_jobs:Creating PipelineJob

INFO:google.cloud.aiplatform.pipeline\_jobs:PipelineJob created. Resource name: projects/267810977261/locations/us-central1/pipelineJobs/automl-tab-beans-training-v2-20211208065510

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/267810977261/locations/us-central1/pipelineJobs/automl-tab-beans-training-v2-20211208065510')

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-v2-20211208065510?project=267810977261>

After navigating to the link given, we see the below results:

#### 1) ML Pipeline

Google Cloud Platform CMPE297Project 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-v2-20211208065510

Runtime Graph 6/6 steps completed Expand Artifacts 90%

```

graph TD
    A[tabulardataset-create] --> B[dataset]
    B --> C[automltabulartrainingjob]
    C --> D[model]
    D --> E[classification-model-eval]
    E --> F[condition-deploy-decision]
    E --> G[metrics]
    E --> H[metrics]
  
```

Pipeline run analysis

SUMMARY NODE INFO

Execution Info **Completed**

VIEW JOB VIEW LOGS

Display name classification-model-eval-metrics

Name classification-model-eval-metrics

Type system.ContainerE

Duration 4 min 14 sec

Started Dec 8, 2021, 12:56:47 AM

Completed Dec 8, 2021, 1:01:02 AM

Input Parameters

Parameter	Type	Value
location	string	central1
project	string	cmpe297project
api_endpoint	string	central1-aiplatform.googleapis.com
thresholds_dict	string	{'auRoc': 0.95}

Output Parameters

Parameter	Type	Value
dep_decision	string	true

## 2) Dataset:

Google Cloud Platform CMPE297Project Search products and resources

Vertex AI

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

Marketplace

automl-beans1638946452

SOURCE ANALYZE

INTEGER 2 (11.76%)  
FLOAT 14 (82.35%)  
STRING 1 (5.88%)

General statistics generated by Dec 07, 2021 11:03 PM GENERATE STATISTICS

Filter Enter property name or value

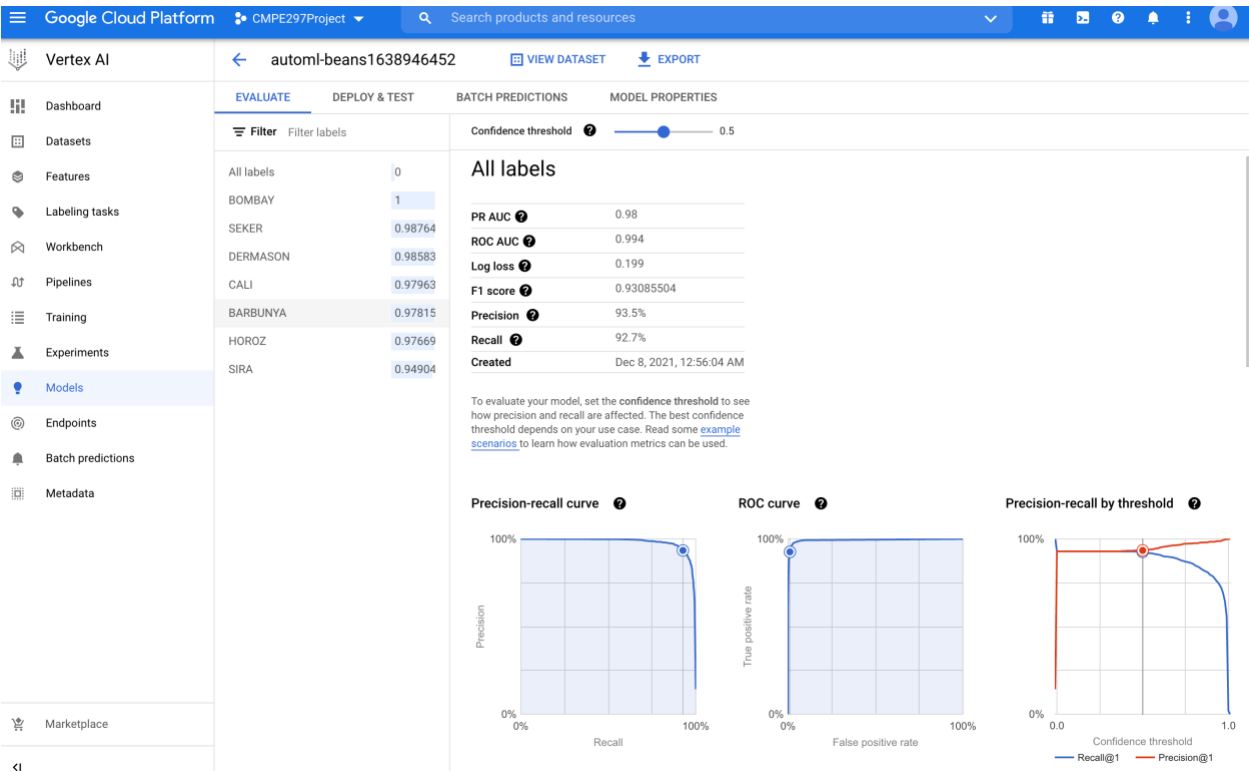
Column name	BigQuery type	BigQuery mode	Missing % (count)	Distinct values
Area	INTEGER	NULLABLE	-	12011
AspectRatio	FLOAT	NULLABLE	-	13543
Class	STRING	NULLABLE	-	7
Compactness	FLOAT	NULLABLE	-	13543
ConvexArea	INTEGER	NULLABLE	-	12066
Eccentricity	FLOAT	NULLABLE	-	13543
EquivDiameter	FLOAT	NULLABLE	-	12011
Extent	FLOAT	NULLABLE	-	13535
MajorAxisLength	FLOAT	NULLABLE	-	13543
MinorAxisLength	FLOAT	NULLABLE	-	13543
Perimeter	FLOAT	NULLABLE	-	13351
roundness	FLOAT	NULLABLE	-	13543
ShapeFactor1	FLOAT	NULLABLE	-	13543
ShapeFactor2	FLOAT	NULLABLE	-	13543
ShapeFactor3	FLOAT	NULLABLE	-	13543
ShapeFactor4	FLOAT	NULLABLE	-	13540
Solidity	FLOAT	NULLABLE	-	13526

Training jobs and models

automl-beans1638946452  
Model type: Tabular

TRAIN NEW MODEL

## 3) Model artifacts:



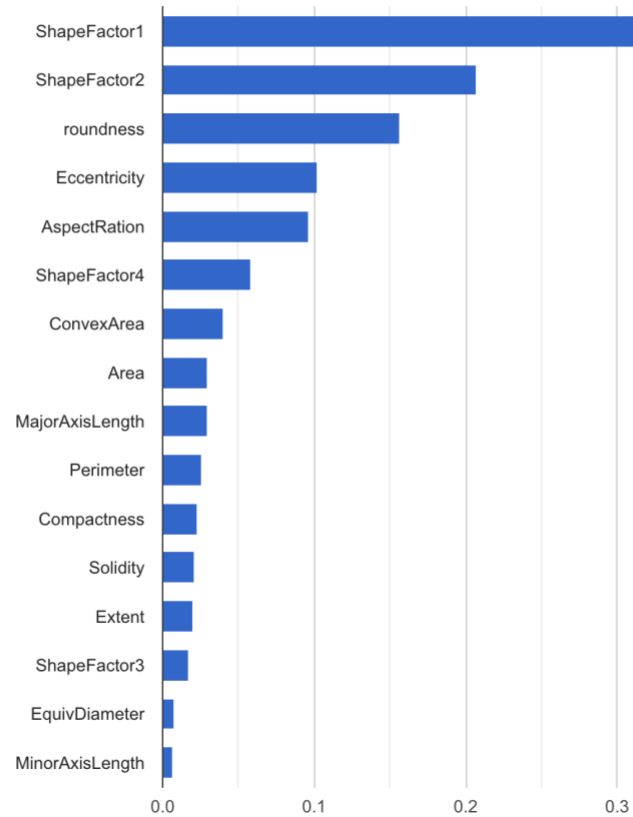
## Confusion matrix

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	DERMASON	SIRA	SEKER	HOROZ	CALI	BARBUNYA	BOMBAY
DERMASON	93%	5%	2%	—	—	—	—	
SIRA	7%	90%	1%	2%	—	1%	—	
SEKER	1%	3%	96%	—	—	—	—	
HOROZ	1%	3%	—	96%	1%	—	—	
CALI	—	1%	—	3%	95%	2%	—	
BARBUNYA	—	4%	1%	—	5%	90%	—	
BOMBAY	—	—	—	—	—	—	100%	

## Feature Importance



We also see that the model is deployed to the endpoint:

Endpoints <span>+ CREATE ENDPOINT</span> <span>REFRESH</span>											
<p>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.</p> <p>To create an endpoint, you need at least one machine learning model. <a href="#">Learn more</a></p> <p>Region: <span>us-central1 (Iowa)</span></p>											
<div>Filter Enter a property name</div>											
<input type="checkbox"/>	Name	ID	Status	Models	Region	Monitoring	Most recent alerts	Last updated ↓	API	Notification	Labels
<input type="checkbox"/>	train-automl-beans	630583112669069312	Active	1	us-central1	Disabled	—	Dec 8, 2021, 1:01:18 AM	Sample request		
<input type="checkbox"/>	cmpe297_project-endpoint	7998472103047200768	Active	1	us-central1	Disabled	—	Dec 5, 2021, 3:16:45 PM	Sample request		

4) All other artifacts:

## Pipeline run analysis

### NODE INFO

## Running



## Running

[VIEW LOGS](#)

Display name	automltabulartrainingjob-run
Name	automltabulartrainingjob-run
Type	system.ContainerExecution
Duration	44 sec
Started	Dec 7, 2021, 10:59:02 PM
Completed	

### Input Parameters

Parameter	Type	Value
project	string	cmpe297/project
column_transformations	string	[{"numeric": { "column_name": "Area"}, {"numeric": { "column_name": "Perimeter"}, {"numeric": { "column_name": "MajorAxisLength"}}, {"numeric": { "column_name": "MinorAxisLength"}}, {"numeric": { "column_name": "AspectRatio"}}, {"numeric": { "column_name": "Eccentricity"}}, {"numeric": { "column_name": "ConvexArea"}}, {"numeric": { "column_name": "EquivalentDiameter"}, {"numeric":

## Pipeline run analysis

SUMMARY	<b>NODE INFO</b>																																	
<p><b>Execution Info</b> <span style="float: right;"> Pending</span></p> <hr/> <table style="width: 100%;"> <tr> <td><b>Display name</b></td> <td>classification-model-eval-metrics</td> </tr> <tr> <td><b>Name</b></td> <td>classification-model-eval-metrics</td> </tr> <tr> <td><b>Type</b></td> <td>container</td> </tr> <tr> <td><b>Duration</b></td> <td>0 sec</td> </tr> <tr> <td><b>Started</b></td> <td></td> </tr> <tr> <td><b>Completed</b></td> <td></td> </tr> </table> <hr/> <p><b>Input Parameters</b></p> <table style="width: 100%;"> <thead> <tr> <th>Parameter</th> <th>Type</th> <th>Value</th> </tr> </thead> <tbody> <tr> <td>location</td> <td>string</td> <td>Cannot fetch parameter value since it's static</td> </tr> <tr> <td>project</td> <td>string</td> <td>Cannot fetch parameter value since it's static</td> </tr> <tr> <td>api_endpoint</td> <td>string</td> <td>Cannot fetch parameter value since it's static</td> </tr> <tr> <td>thresholds_dict_str</td> <td>string</td> <td>Cannot fetch parameter value since it's static</td> </tr> </tbody> </table> <hr/> <p><b>Output Parameters</b></p> <table style="width: 100%;"> <thead> <tr> <th>Parameter</th> <th>Type</th> <th>Value</th> </tr> </thead> <tbody> <tr> <td>dep_decision</td> <td>string</td> <td>&lt;undefined_till_runtime&gt;</td> </tr> </tbody> </table>		<b>Display name</b>	classification-model-eval-metrics	<b>Name</b>	classification-model-eval-metrics	<b>Type</b>	container	<b>Duration</b>	0 sec	<b>Started</b>		<b>Completed</b>		Parameter	Type	Value	location	string	Cannot fetch parameter value since it's static	project	string	Cannot fetch parameter value since it's static	api_endpoint	string	Cannot fetch parameter value since it's static	thresholds_dict_str	string	Cannot fetch parameter value since it's static	Parameter	Type	Value	dep_decision	string	<undefined_till_runtime>
<b>Display name</b>	classification-model-eval-metrics																																	
<b>Name</b>	classification-model-eval-metrics																																	
<b>Type</b>	container																																	
<b>Duration</b>	0 sec																																	
<b>Started</b>																																		
<b>Completed</b>																																		
Parameter	Type	Value																																
location	string	Cannot fetch parameter value since it's static																																
project	string	Cannot fetch parameter value since it's static																																
api_endpoint	string	Cannot fetch parameter value since it's static																																
thresholds_dict_str	string	Cannot fetch parameter value since it's static																																
Parameter	Type	Value																																
dep_decision	string	<undefined_till_runtime>																																

# Pipeline run analysis

SUMMARY    **NODE INFO**

## Execution Info

 Pending

Display name	condition-deploy-decision-1
Name	condition-deploy-decision-1
Type	
Duration	0 sec
Started	
Completed	

## Input Parameters

Parameter	Type	Value
pipelineparam--classification-model-eval-metrics-dep_decision	string	<undefined_till_runtime>
pipelineparam--gcp_region	string	Cannot fetch parameter value since it's static
pipelineparam--project	string	Cannot fetch parameter value since it's static

# Pipeline run analysis

**SUMMARY**    NODE INFO

## Basic info

Duration	6 min 56 sec
Started	Dec 7, 2021, 10:55:17 PM
Completed	
Run name	automl-tab-beans-training-v2-20211208065510
Pipeline name	automl-tab-beans-training-v2
Runtime environment	Serverless
Region	us-central1
Service account	267810977261-compute@developer.gserviceaccount.com
Debugging info	<a href="#">View pipeline proto</a>

## Run Parameters

Pipeline parameter values used for this run

Parameter	Type	Value
gcp_region	string	us-central1
bq_source	string	bq://aju-dev-demos.beans.beans1
thresholds_dict_str	string	{"auRoc": 0.95}
project	string	cmpe297project
display_name	string	automl-beans1638946452
api_endpoint	string	us-central1-aiplatform.googleapis.com



# Pipeline run analysis

SUMMARY

NODE INFO

## Basic info

Duration	7 min 16 sec
Started	Dec 7, 2021, 10:55:17 PM
Completed	
Run name	automl-tab-beans-training-v2-20211208065510
Pipeline name	automl-tab-beans-training-v2
Runtime environment	Serverless
Region	us-central1
Service account	267810977261-compute@developer.gserviceaccount.com
Debugging info	<a href="#">View pipeline proto</a>

## Run Parameters

Pipeline parameter values used for this run

Parameter	Type	Value
gcp_region	string	us-central1
bq_source	string	bq://aju-dev-demos.beans.beans1
thresholds_dict_str	string	{"auRoc": 0.95}
project	string	cmpe297project
display_name	string	automl-beans1638946452
api_endpoint	string	us-central1-aiplatform.googleapis.com