# Intro to Vertex Pipelines

This document shows the details of creating and running ML Pipelines

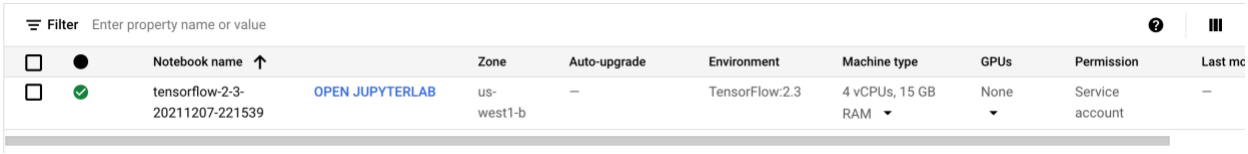
1. Setting the project and enabling APIs required to build vertex pipelines

```
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

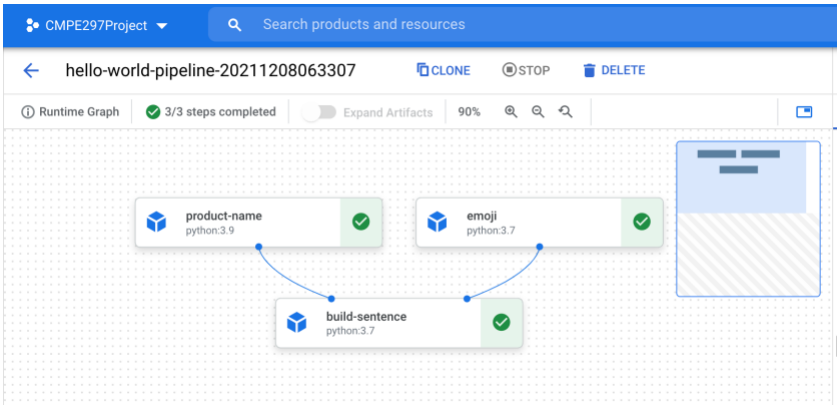
```
arpithagurumurthy_ml@cloudshell:~ (cmpe297project)$ BUCKET_NAME=gs://$GOOGLE_CLOUD_PROJECT-bucket
arpithagurumurthy_ml@cloudshell:~ (cmpe297project)$ gsutil mb -l us-central1 $BUCKET_NAME
Creating gs://cmpe297project-bucket/...
arpithagurumurthy_ml@cloudshell:~ (cmpe297project)$ gcloud projects describe $GOOGLE_CLOUD_PROJECT > project-info.txt
arpithagurumurthy_ml@cloudshell:~ (cmpe297project)$ PROJECT_NUM=$(cat project-info.txt | sed -nre 's:.*projectNumber\: (.*):\1:p')
arpithagurumurthy_ml@cloudshell:~ (cmpe297project)$ 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.objectAdmi
n
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@gcf-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: BwXSnFDcnS0=
version: 1
arpithagurumurthy_ml@cloudshell:~ (cmpe297project)$
```

3. Creating a Vertex AI Workbench instance

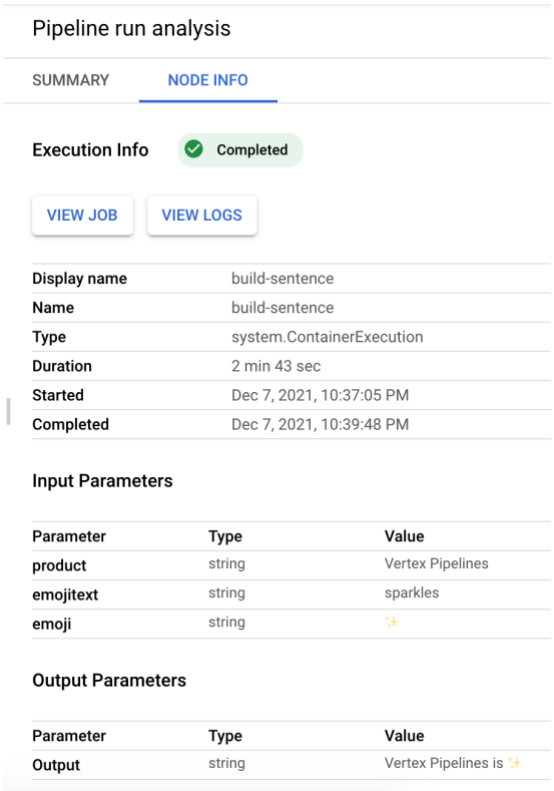



**Attaching the Python notebook for the pipeline in github

## Pipeline after completion




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



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:

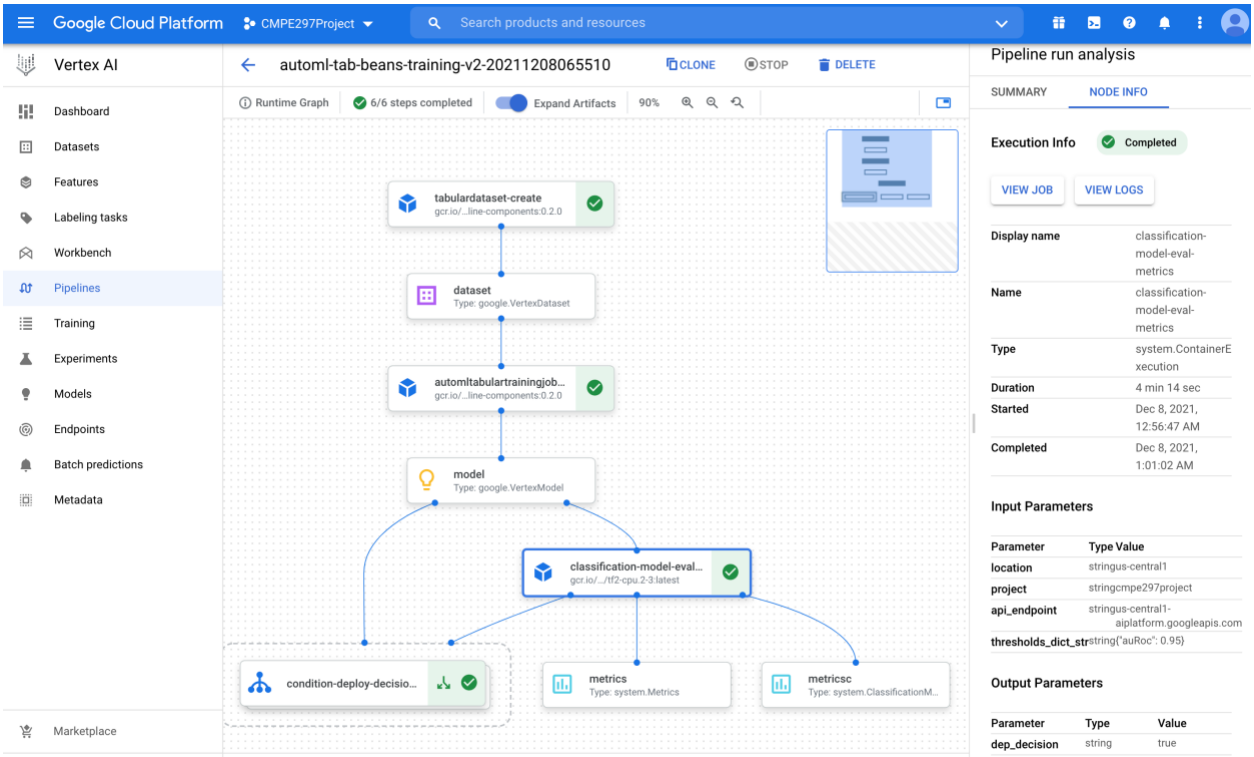
- Create a dataset in Vertex AI
- Train a tabular classification model with AutoML
- Get evaluation metrics on this model
- Deploy the model to an endpoint

**Attaching the Python notebook for the pipeline in github

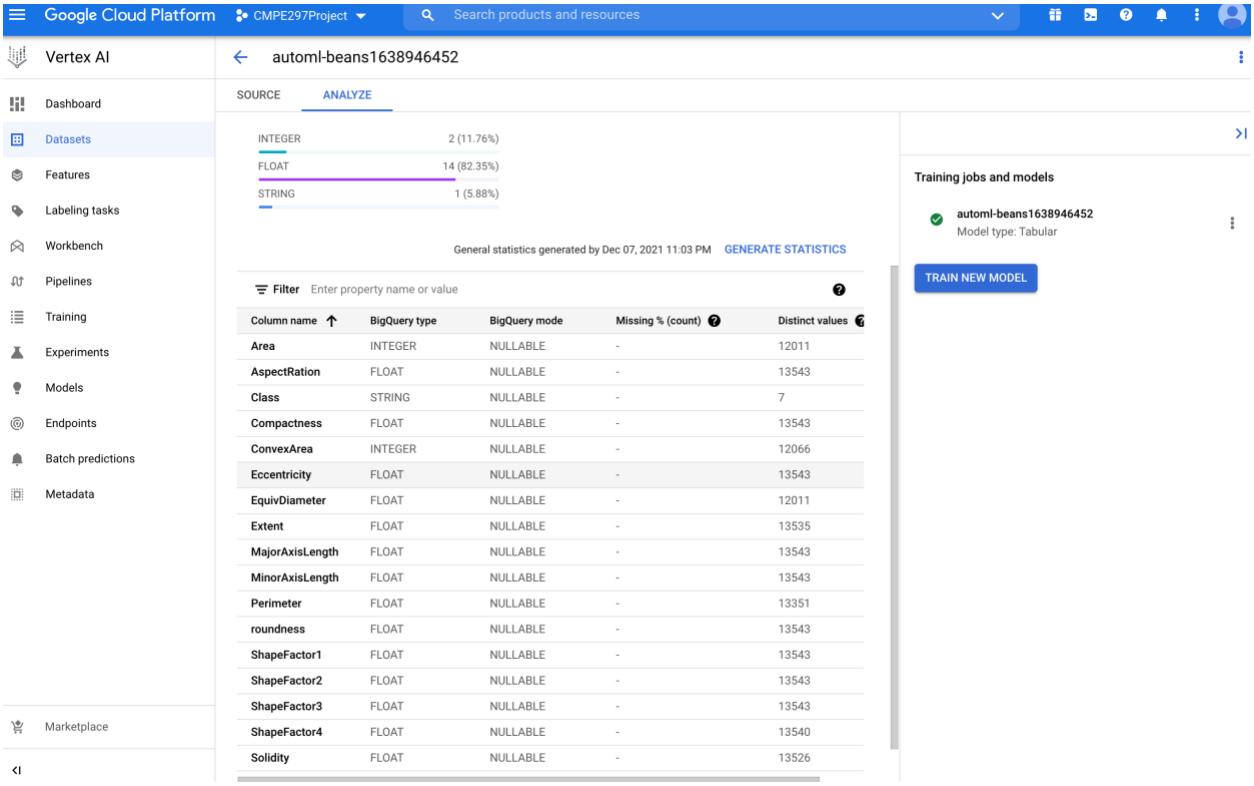
Output:
INFO:google.cloud.aiplatform.pipeline_jobs:Creating PipelineJob
INFO:google.cloud.aiplatform.pipeline_jobs:PipelineJob created. Resource name: projects/267810977261/locations/us-central1/pipelineJobs/automl-tab-beans-training-v2-20211208065510
INFO:google.cloud.aiplatform.pipeline_jobs:To use 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:

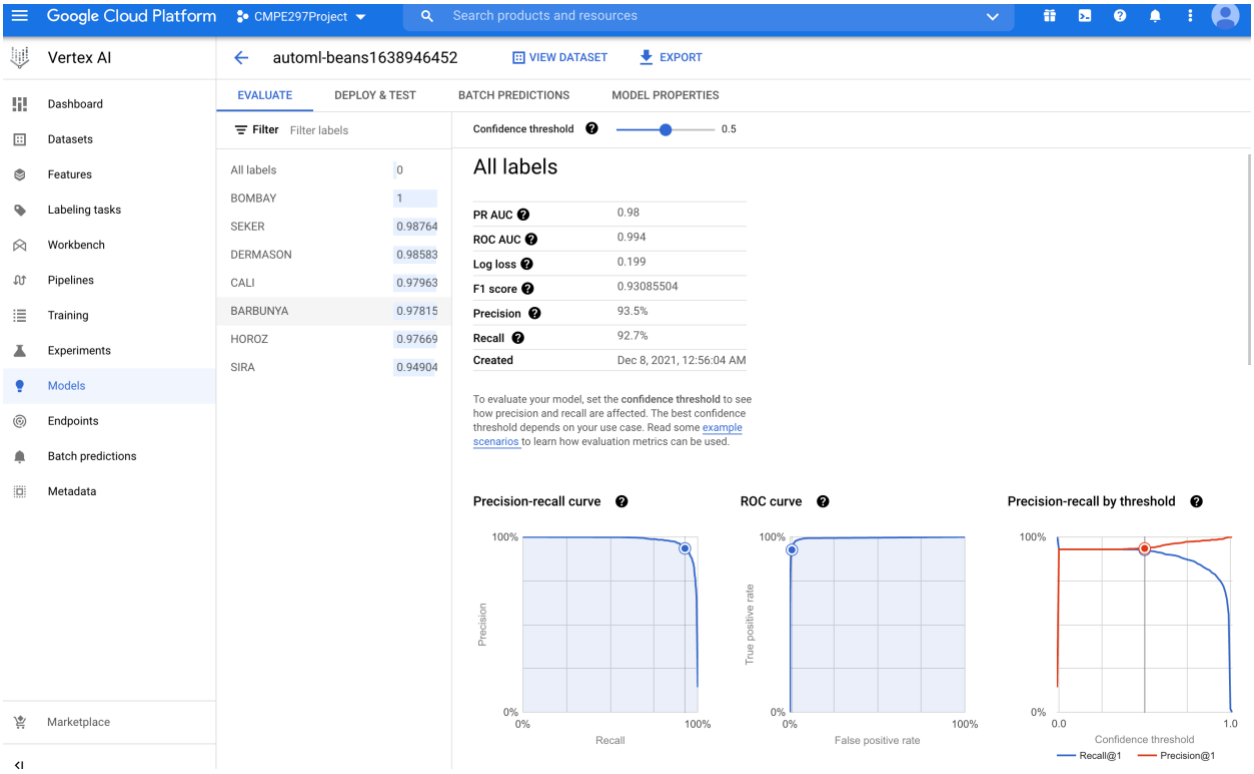
1) ML Pipeline




2) Dataset:




3) Model artifacts:

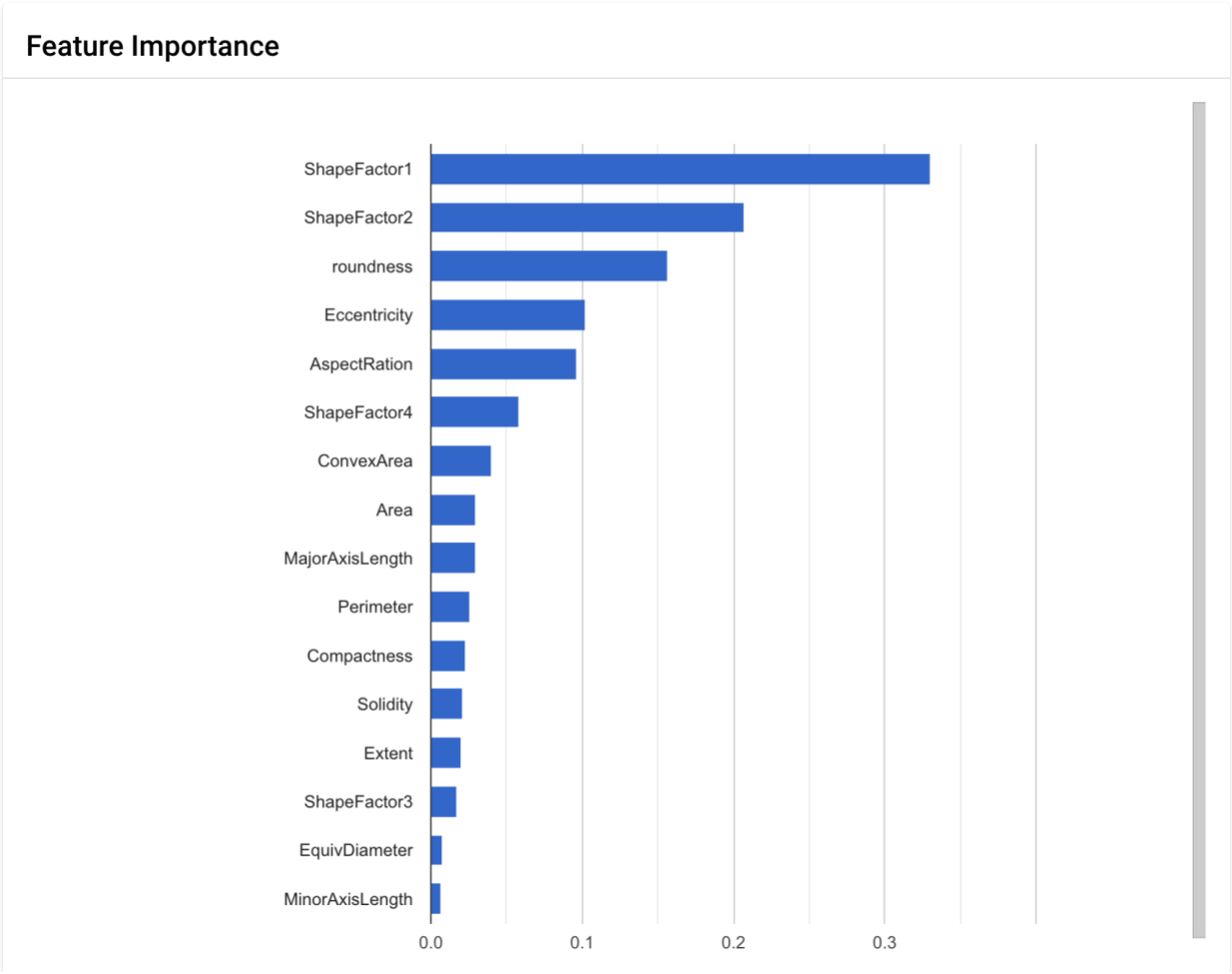



## Confusion matrix

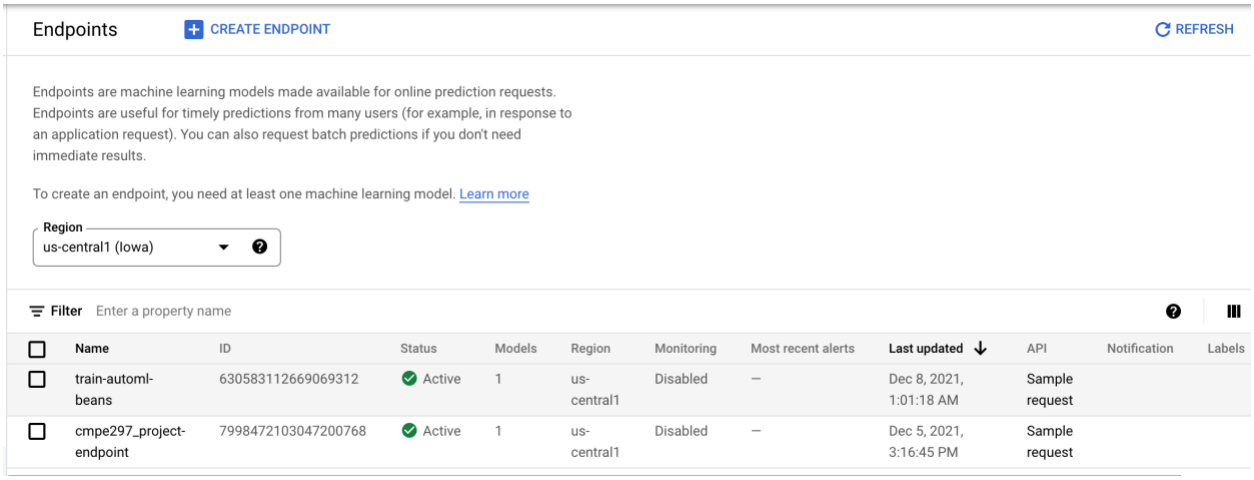
Item counts

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% |




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




4) All other artifacts:

Pipeline run analysis

SUMMARY NODE INFO

**Execution Info** Running

VIEW JOB 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 | cmpe297project |
| **column_transformations** | string | [{"numeric": {"column_name": "Area"}}, {"numeric": {"column_name": "Perimeter"}}, {"numeric": {"column_name": "MajorAxisLength"}}, {"numeric": {"column_name": "MinorAxisLength"}}, {"numeric": {"column_name": "AspectRation"}}, {"numeric": {"column_name": "Eccentricity"}}, {"numeric": {"column_name": "ConvexArea"}}, {"numeric": {"column_name": "EquivDiameter"}}, |

Pipeline run analysis

SUMMARY NODE INFO

**Execution Info** Pending

| | |
|---|---|
| **Display name** | classification-model-eval-metrics |
| **Name** | classification-model-eval-metrics |
| **Type** | container |
| **Duration** | 0 sec |
| **Started** | |
| **Completed** | |

**Input Parameters**

| 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 |

**Output Parameters**

| 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** | View pipeline proto |

### 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** | View pipeline proto |

## 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 |