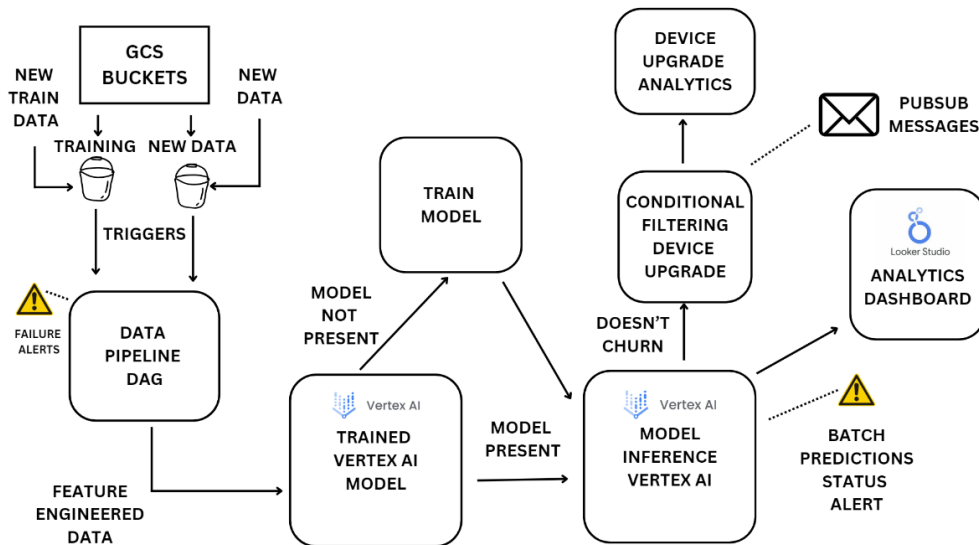


Telecom Customer Churn and Mobile Device Upgrade Prediction

Project Workflow and Screenshots



Business Use Case: Telecom companies have a pressing need to analyze dynamic customer behavior in order to send timely offers and prevent customer churn. This project addresses the need by designing a GCP based end-to-end pipeline for predicting customer churn and filtering customers who would upgrade their device to send notifications.

As the telecom companies receive customer data in new **batches**, this pipeline gets triggered to perform churn prediction, device upgrade estimation, and real-time customer analytics on interactive dashboards. We designed an interconnected DAG system with triggers to orchestrate the flow as new data gets ingested.

GCP Project Service Account Permissions:

	mlops-project-service-account@axial-rigging-438817-h4.iam.gserviceaccount.com	mlops-project-service-account	Cloud Build Logging Service Agent
			Composer Worker
			Logs Configuration Writer
			Logs Viewer
			Monitoring Editor
			Monitoring Viewer
			Owner
			Storage Admin
			Storage Object Viewer
			10251/10303 excess permis

GCS Buckets for Data:

Google Cloud

MLOpsProject

Search (/) for resources, docs, products, and more

Search

Cloud Storage

Buckets

CREATE

REFRESH

GO TO PATH

LEARN

Overview

Buckets

Monitoring

Settings

Filter

Filter buckets

Name	Created	Location type	Location	Default storage class	Last
additional	Dec 4, 2024, 6:51:43 PM	Multi-region	us	Standard	Dec
batch_predict_test_processed	Nov 11, 2024, 8:56:13 PM	Region	us-central1	Standard	Nov
cloud-ai-platform-922b6460-cadb-450e...	Nov 9, 2024, 5:37:35 PM	Region	us-central1	Regional	Nov
data-source-telecom-customers	Oct 22, 2024, 1:25:19 PM	Region	us-east1	Standard	Nov
data_slices_for_bias_detection	Nov 13, 2024, 4:55:59 PM	Region	us-central1	Standard	Nov
device-upgrade-prediction-dvc	Oct 22, 2024, 1:21:21 PM	Region	us-east1	Standard	Oct
gcf-v2-sources-902448813225-us-centra...	Nov 29, 2024, 1:51:59 PM	Region	us-central1	Standard	Nov
gcf-v2-uploads-902448813225-us-centra...	Nov 29, 2024, 1:51:57 PM	Region	us-central1	Standard	Nov
holdout_batches	Nov 22, 2024, 12:47:17 AM	Region	us-central1	Standard	Nov
us-central1-automi-composer-a048d1fd...	Nov 14, 2024, 11:51:26 AM	Region	us-central1	Standard	Nov
us-central1-churn-predictio-95fd53a7-bu...	Nov 20, 2024, 5:10:13 PM	Region	us-central1	Standard	Nov
us-central1-new-automi-comp-9ffd484f...	Dec 3, 2024, 5:02:06 PM	Region	us-central1	Standard	Dec
us-central1-telcom-churn-97999668b...	Nov 12, 2024, 9:51:49 PM	Region	us-central1	Standard	Nov

Buckets

>

data-source-telecom-customers

>

data

>

raw_data

CREATE FOLDER

UPLOAD

TRANSFER DATA

OTHER SERVICES

Filter by name prefix only

Filter

Filter objects and folders

Show

Live objects only

Name	Size	Type	Created	Storage class	Last modified	Public access	Version
new_data/	—	Folder	—	—	—	—	—
test/	—	Folder	—	—	—	—	—
train/	—	Folder	—	—	—	—	—

DVC integration for data version control:

device-upgrade-prediction-dvc

Location

Storage class

Public access

Protection

Hierarchical namespace

us-east1 (South Carolina)

Standard

Not public

Soft Delete

Enabled

OBJECTS

CONFIGURATION

PERMISSIONS

PROTECTION

LIFECYCLE

OBSERVABILITY

INVENTORY REPORTS

OPERATIONS

Buckets

>

device-upgrade-prediction-dvc

CREATE FOLDER

UPLOAD

TRANSFER DATA

OTHER SERVICES

Filter by name prefix only

Filter

Filter objects and folders

Show

Live objects only

Name	Size	Type	Created	Storage class	Last modified	Public access	Version history	Encrypt
files/	—	Folder	—	—	—	—	—	—

DVC logs cache files stored in a bucket.






Python modules loaded to GCS:

Custom made python modules that are used by DAG tasks. DAG access these modules from the bucket.

Buckets > data-source-telecom-customers > python_modules

CREATE FOLDER UPLOAD TRANSFER DATA OTHER SERVICES

Filter by name prefix only Filter Filter objects and folders

<input type="checkbox"/>	Name	Size	Type	Created
<input type="checkbox"/>	 axial-rigging-438817-h4-6e7c2d5...	2.4 KB	application/json	Nov 25, 2024, 7:14:04 P
<input type="checkbox"/>	 data_loader.py	1.1 KB	application/octet-stream	Nov 15, 2024, 6:25:50 P
<input type="checkbox"/>	 feature_engineering.py	3.6 KB	application/octet-stream	Nov 24, 2024, 9:10:44 P
<input type="checkbox"/>	 preprocessing.py	8.6 KB	application/octet-stream	Nov 24, 2024, 9:10:44 P
<input type="checkbox"/>	 upload_to_gcs.py	647 B	application/octet-stream	Nov 15, 2024, 6:27:02 P

DAGs in Composer environment:















Composer Environment d... OPEN AIRFLOW UI OPEN DAGS FOLDER SAVE SNAPSHOT LOAD SNAPSHOT

automl-composer This environment is running

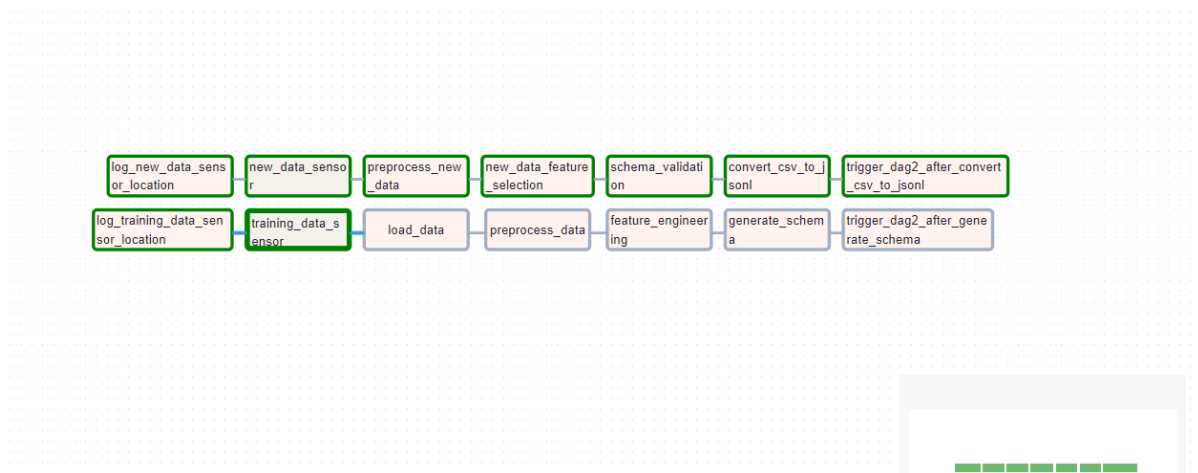
MONITORING LOGS DAGS ENVIRONMENT CONFIGURATION AIRFLOW CONFIGURATION OVERRIDES ENVIRONMENT VARIABLES LABELS PYPI PACKAGES

Filter Filter DAGs 1 hour 6 hours 12 hours 1 day 2 days 4 days 7 days 14 days 30 days

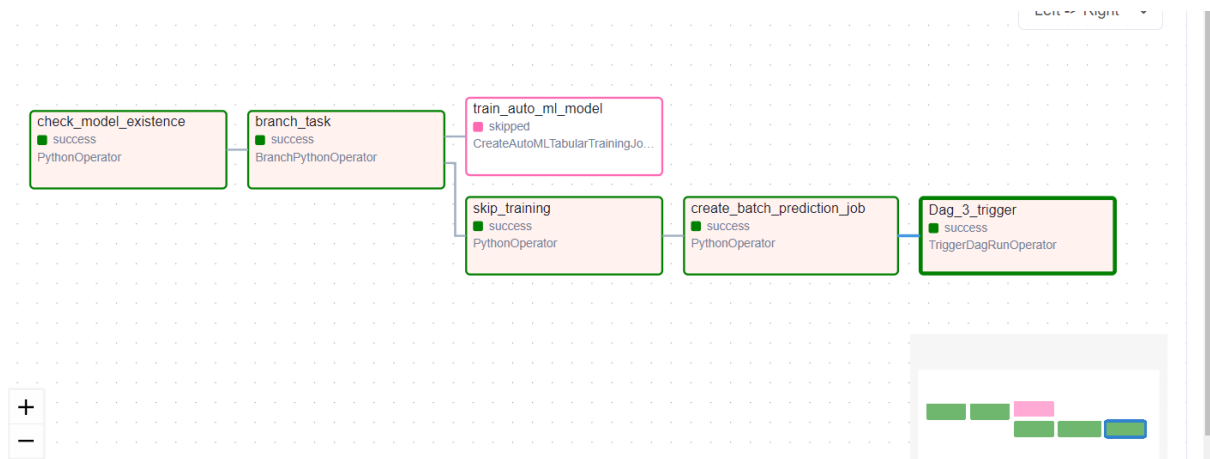
DAG id	State	Description	Schedule interval	Last completed run	Active runs	Successful runs (1h)	Failed runs (1h)
airflow_monitoring	Active	liveness...	* / 10 * * * *	Just now	0	6	
bigquery_to_pubsub	Active	Fetch...		4 hours ago	0	0	
telecom_dag_new	Active	Datapipe...		1 day ago	0	0	
vertex_ai_churn_model_training	Active			4 hours ago	0	0	

<div>  Airflow </div> <div> DAGs Cluster Activity Datasets Browse Admin Docs Composer </div> <div>00:31 UTC V-</div>									
automl-composer									
<div> <div> All 4 Active 4 Paused 0 Running 0 Failed 1 </div> <div> Filter DAGs by tag Search DAGs </div> <div> Auto-refresh </div> </div>									
DAG	Owner	Runs	Schedule	Last Run	Next Run	Recent Tasks			
 airflow_monitoring	airflow	 2229	 */10 * * * *	2024-12-05, 00:20:00	2024-12-05, 00:30:00				
 bigquery_to_pubsub	airflow	 3	None	2024-12-04, 19:40:47					
 telecom_dag_new	airflow	 3	None	2024-12-03, 23:36:20					
 vertex_ai_churn_model_training	airflow	 10	None	2024-12-04, 19:26:36					

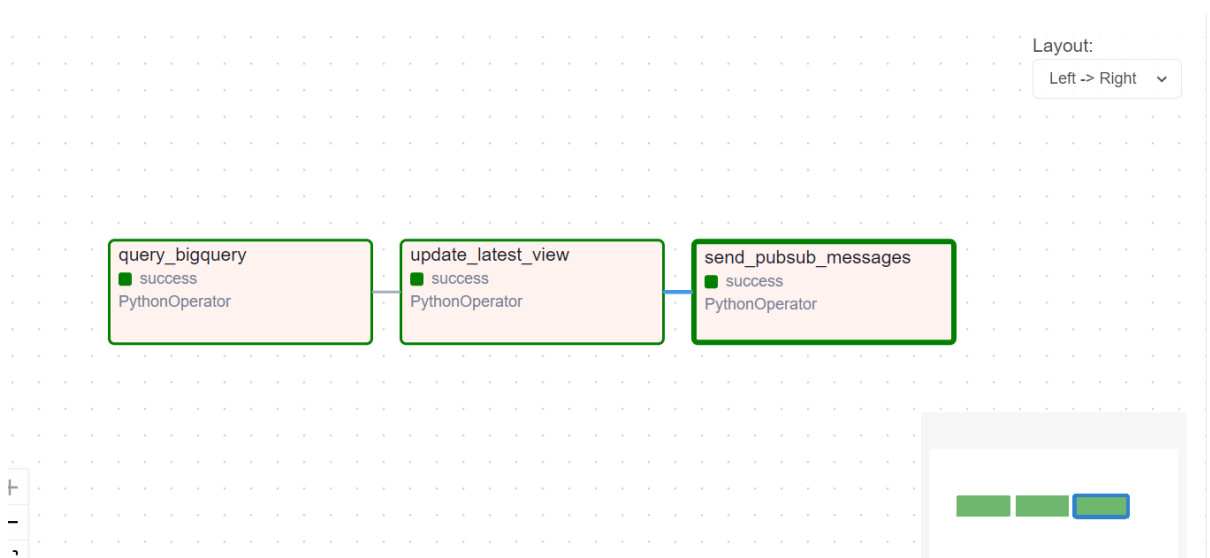
Data Pipeline DAG – gets triggered with incoming new data



Vertex AI DAG - gets triggered when data DAG runs successfully; makes a call to Vertex AI model (Inference if model is present, training if model is not present already)



Front end DAG - gets triggered when vertex AI DAG runs successfully; Stores predictions in Big Query, updates Looker dashboard, filters customers likely to upgrade and sends messages Pub/Sub



Vertex AI model:

Vertex AI

TOOLS

Dashboard

Model Garden

Pipelines

NOTEBOOKS

Colab Enterprise

Workbench

VERTEX AI STUDIO

Overview

Freeform

Marketplace

Training

TRAIN NEW MODEL

REFRESH

LEARN

TRAINING PIPELINES

CUSTOM JOBS

HYPERPARAMETER TUNING JOBS

NAS JOBS

PERSISTENT RESOURCES

Region

us-central1 (Iowa)

Filter

Enter a property name

Name	ID	Status	Job type	Model type	Duration	Last updated	Created
churn_model_2	6764848472186159104	Finished	Training pipeline	Tabular classification	3 hr 20 min	Nov 20, 2024, 10:20:02 PM	Nov 20, 2024, 6:59:16 PM
churn_model_1	3122952600178130944	Finished	Training pipeline	Tabular classification	2 hr 7 min	Nov 14, 2024, 4:08:23 PM	Nov 14, 2024, 2:00:56 PM
churn_model	8002862281178218496	Finished	Training pipeline	Tabular classification	2 hr 3 min	Nov 11, 2024, 7:31:45 PM	Nov 11, 2024, 5:28:40 PM

Latest and best model version – churn_model_2.

Notifications: Email alerts received after training completion and every batch prediction completion.

On Sat, Nov 9, 2024 at 9:26 PM Vertex AI <noreply-vertexai@google.com> wrote:

Hello Vertex AI Customer,

Vertex AI finished training model "temp_churn".

Additional Details:

Operation State: Succeeded

Resource Name:

projects/902448813225/locations/us-central1/trainingPipelines/6226524830447960064

To continue your progress, go back to your training pipeline using <https://console.cloud.google.com/vertex-ai/models?authuser=1&hl=en&project=axial-rigging-438817-h4>

Sincerely,

The Google Cloud AI Team

Vertex AI finished batch prediction using model "bias_detection_currentequip_slice1"

Vertex AI <noreply-vertexai@google.com> to me

Thu, Nov 14, 5:51 PM

Hello Vertex AI Customer,

Vertex AI finished batch prediction using model "bias_detection_currentequip_slice1".

Additional Details:

Operation State: Succeeded

Resource Name:

projects/902448813225/locations/us-central1/batchPredictionJobs/9134017853601611776

To continue your progress, go back to your predictions using <https://console.cloud.google.com/vertex-ai/models/locations/us-central1/models/2993104846782988288/versions/1/batch-predictions?referrer=search&project=axial-rigging-438817-h4>

Sincerely,

The Google Cloud AI Team

Vertex AI model inferences stored on Big Query for every batch:

```
# Task: Create batch prediction job
def create_batch_prediction_job(**kwargs):
    log = get_task_logger("create_batch_prediction_job")
    try:
        model_name = kwargs['ti'].xcom_pull(task_ids='check_model_existence', key='model_name')
        if not model_name:
            raise ValueError("No model found for batch prediction.")

        log.info("Initializing Vertex AI platform for batch prediction...")
        aiplatform.init(project=PROJECT_ID, location=LOCATION)

        gcs_input = f"gs://{BUCKET_NAME}/latest_best_features_for_churn.jsonl"
        bigquery_output_prefix = "axial-rigging-438817-h4.Big_query_batch_prediction"

        log.info("Creating batch prediction job...")
        batch_prediction_job = aiplatform.BatchPredictionJob.create(
            job_display_name="batch_prediction_job",
            model_name=model_name,
            gcs_source=gcs_input,
            predictions_format="bigquery",
            starting_replica_count=10,
            max_replica_count=20,
            machine_type="c2-standard-30",
            bigquery_destination_prefix=bigquery_output_prefix
```

The screenshot shows the Google Cloud BigQuery Studio interface. The left sidebar contains navigation options like Analysis, Data transfers, Scheduled queries, Analytics Hub, Dataform, Partner Center, Orchestration, Migration, Assessment, SQL translation, and Release Notes. The main area displays a table named 'predictions_2...' with columns: Row, Age, UnansweredCalls, UniqueSubs, predicted_Churn.classes, and pred_scores. The table contains 7 rows of data. Below the table, there is a 'Job history' section with a 'REFRESH' button.

Row	Age	UnansweredCalls	UniqueSubs	predicted_Churn.classes	pred_scores
1		42.0	2	0	0.87005174...
2		7.3	1	1	0.12994828...
3		19.7	2	1	0.86864614...
4		23.0	1	0	0.13135384...
5		16.3	1	0	0.73005568...
6		37.7	3	1	0.26994436...
7		11.7	2	1	0.77415192...

Model Performance:

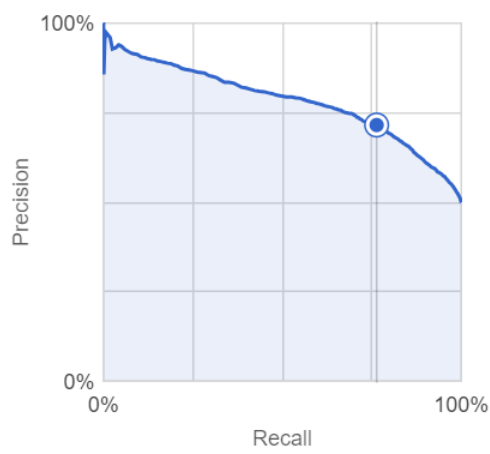
Evaluation details

Confidence threshold ? 0.5

All labels

PR AUC ?	0.782
ROC AUC ?	0.793
Log loss ?	0.551
Micro-average F1 ?	0.72627234
Macro-average F1 ?	0.5111234
Micro-average precision ?	72.6%
Micro-average recall ?	72.6%



Precision-recall curve ?



ROC curve



Confusion matrix









 Item counts 

A confusion matrix shows how the model classified each label in the evaluation dataset. The blue, bold cells indicate a correct prediction. A data item is moved to the dropped column if it does not meet the confidence threshold for any label.

True label	Predicted label		
	0	1	Dropped
0	97%	3%	0%
1	89%	11%	0%

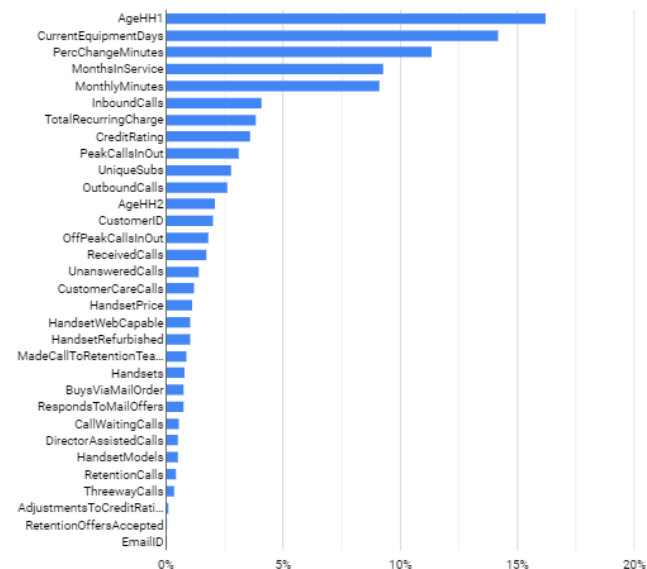
Feature Sensitivity Analysis:

Explanations stored in big query

	predictions_2...	 QUERY	 SHARE	 COPY	 SNAPSHOT	 DELETE	 EXPORT	 REFRESH
SCHEMA	DETAILS	PREVIEW	TABLE EXPLORER	PREVIEW	INSIGHTS	LINEAGE	DATA PROFILE	DATA QUALITY
<input type="checkbox"/>	RetentionOffers		STRING	NULLABLE	-	-	-	-
<input type="checkbox"/>	RetentionOffersAccepted		STRING	NULLABLE	-	-	-	-
<input type="checkbox"/>	ThreewayCalls		STRING	NULLABLE	-	-	-	-
<input type="checkbox"/>	TotalRecurringCharge		STRING	NULLABLE	-	-	-	-
<input type="checkbox"/>	UnansweredCalls		STRING	NULLABLE	-	-	-	-
<input type="checkbox"/>	UniqueSubs		STRING	NULLABLE	-	-	-	-
<input type="checkbox"/>	▼ explanation		RECORD	NULLABLE	-	-	-	-
<input type="checkbox"/>	▼ attributions		RECORD	REPEATED	-	-	-	-
<input type="checkbox"/>	► featureAttributions		RECORD	NULLABLE	-	-	-	-
<input type="checkbox"/>	outputDisplayName		STRING	NULLABLE	-	-	-	-
<input type="checkbox"/>	► predicted_Churn		RECORD	NULLABLE	-	-	-	-

Feature importance

Model feature attribution tells you how important each feature is when making a prediction. Attribution values are expressed as a percentage; the higher the percentage, the more strongly that feature impacts a prediction on average. Model feature attribution is expressed using the Sampled Shapley method. [Learn more](#)



Bias Detection:

Bias detection performed on the most relevant column 'CurrentEquipmentDays'. Data sliced into 3 slices, predictions stored in big query and queried to analyze the model's performance on different slices. The model performance is similar for all slices so no bias was detected.

Explorer

Search BigQuery resources

Viewing resources.

SHOW STARRED ONLY

axial-rigging-438817-h4

Queries

Shared queries

bias_detection_query

Notebooks

Data canvases

Data preparations

Workflows

External connections

SUMMARY

ACTIVITY

Dec 1, 2024

Created - Nithya Ala

7:30 PM

prediction... 066

Untitled query

Untitled query

latest_pr... lew

bias_det... ery

bias_detection_query

RUN

SAVE QUERY

DOWNLOAD

SHARE

SCHEDULE

OPEN IN

MORE

WITH RankedClasses AS (
SELECT
Churn,
class,
score,
ROW_NUMBER() OVER (PARTITION BY Churn ORDER BY score DESC) AS rank
FROM (
SELECT
Churn,
class,
score
FROM
axial-rigging-438817-h4.big_query.bias_detection_predictions_predictions_2024_11_14T15:44:14_3227_581

Query results

SAVE RESULTS

EXPLORE DATA

JOB INFORMATION

RESULTS

CHART

JSON

EXECUTION DETAILS

EXECUTION GRAPH

Row

percentage_predictions

1

50.0

Results per page: 50

1 - 1 of 1

Job history

REFRESH

Model Registry:

Vertex AI

TOOLS

Dashboard

Model Garden

Pipelines

NOTEBOOKS

Colab Enterprise

Workbench

VERTEX AI STUDIO

Overview

Freeform

Marketplace

Model Registry

CREATE

IMPORT

REFRESH

LEARN

Models are built from your datasets or unmanaged data sources. There are many different types of machine learning models available on Vertex AI, depending on your use case and level of experience with machine learning. [Learn more](#)

Region
us-central1 (Iowa)

Filter Enter a property name

Name	Default version	Deployment status	Description	Type	Source	Updated	Labels
churn_model_2	1	—	—	Tabular	AutoML training	Nov 20, 2024, 10:20:03 PM	—
churn_model_1	1	—	—	Tabular	AutoML training	Nov 14, 2024, 4:08:24 PM	—
churn_model	1	—	—	Tabular	AutoML training	Nov 11, 2024, 7:31:45 PM	—

Logging and Alerts:

Logs Explorer

Query library

Share link

Preferences

Dec 1, 3:52 PM - Dec 1, 4:47 PM

EST

Run query

Learn

Project logs

Search all fields

All resources

All log names

Error

Correlate by

Log fields

Search fields and values

RESOURCE TYPE

Cloud Composer Environment

36

SEVERITY

Error

Clear x

Timeline

Dec 1, 3:52 PM

Dec 1, 4:48 PM

36 results

Actions

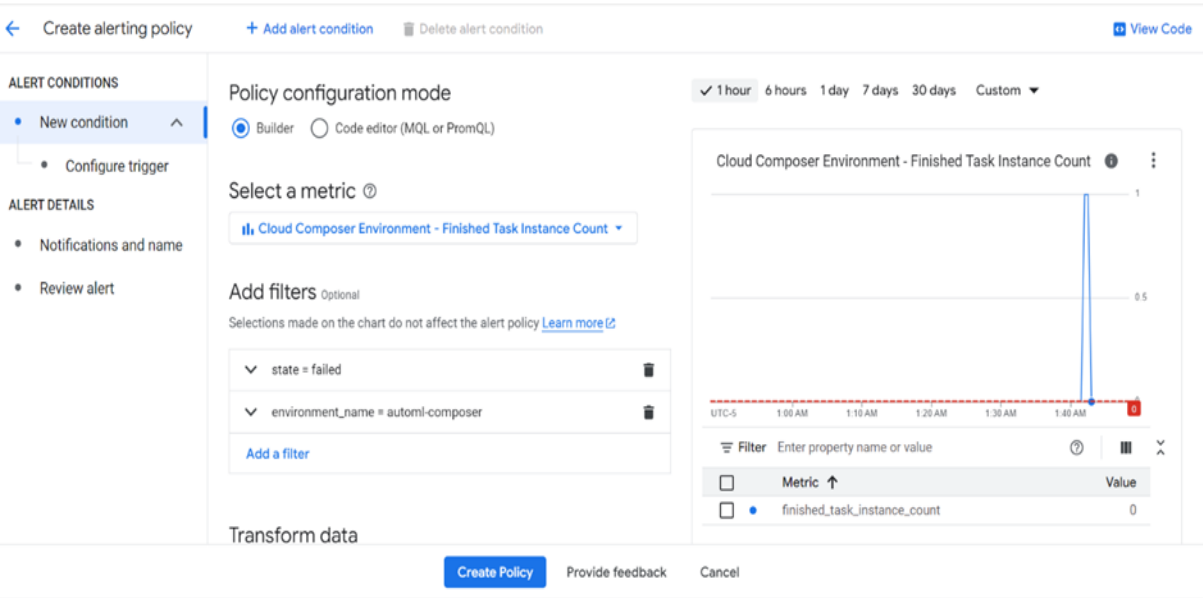
SEVERITY	TIME	SUMMARY
Error	2024-12-01 16:42:51.888	Please set env variable SPARK_VERSION
Error	2024-12-01 16:42:52.602	Please set env variable SPARK_VERSION
Error	2024-12-01 16:42:55.000	Column CustomerCareCalls has values above the maximum allowed.
Error	2024-12-01 16:42:55.000	Validation Error: Column CustomerCareCalls has values above the maximum allowed.
Error	2024-12-01 16:42:55.014	Task failed with exception\nTraceback (most recent call last):\n File "/opt/python3.11/lib/python-
Error	2024-12-01 16:42:55.084	Failed to execute job 3708 for task schema_validation (Column CustomerCareCalls has values above t-
Error	2024-12-01 16:43:02.971	Please set env variable SPARK_VERSION

To view newer entries: Extend time by: 1 minute Edit time

Utilizing log explorer for analyzing logs.

Alert policies set up to monitor task failures and DAG failures.

Email and SMS alerts set up.



Observability Monitoring	Policies	Create policy
Overview	Filter	Filter policies
Dashboards	Display Name	Type
Explore	Error_monitoring	Logs
Metrics explorer	failure_of_DAG	Logs
Logs explorer	task_failure_DAG	Logs
Log analytics		
Trace explorer		
Detect		
Alerting		
Error reporting		

failure_of_DAG

ConditionsSeverity

Policy violates when ANY condition is met

Error

Configurations

Log query

resource.type="cloud_composer_environment" resource.labels.location="us-central1" resource.labels.environment_name="automl-composer" log_name="projects/axial-rigging-438817-h4/logs/airflow-scheduler" severity=ERROR

Notification rate limit

One notification per 5 minutes

Incident autoclose duration

1 day

Logs

SeverityDefault

Filter Search all fields and values

?

↺

↻

SEVERITY	TIMESTAMP	SUMMARY
> <div>Error</div>	2024-11-29 11:12:14.527 EST	airflow-scheduler Marking run <DagRun bigquery_to_pubsub @ 2024-11-29 16:07:07.814909+00:00: manual__2024-11-29T16:07:07.814909+00:00, state:running, queued_at: 202...
> <div>Error</div>	2024-11-29 11:39:06.829 EST	airflow-scheduler Marking run <DagRun bigquery_to_pubsub @ 2024-11-29 16:33:49.737569+00:00: manual__2024-11-29T16:33:49.737569+00:00, state:running, queued_at: 202...
> <div>Error</div>	2024-11-29 12:11:06.733 EST	airflow-scheduler Marking run <DagRun bigquery_to_pubsub @ 2024-11-29 17:05:57.776741+00:00: manual__2024-11-29T17:05:57.776741+00:00, state:running, queued_at: 202...
> <div>Error</div>	2024-12-01 01:47:06.835 EST	airflow-scheduler Marking run <DagRun telecom_dag_new @ 2024-12-01 06:41:33.708044+00:00: manual__2024-12-01T06:41:33.708044+00:00, state:running, queued_at: 2024-1...
> <div>Error</div>	2024-12-01 02:05:45.052 EST	airflow-scheduler Marking run <DagRun telecom_dag_new @ 2024-12-01 07:00:13.094020+00:00: manual__2024-12-01T07:00:13.094020+00:00, state:running, queued_at: 2024-1...
> <div>Error</div>	2024-12-01 15:31:11.861 EST	airflow-scheduler Marking run <DagRun telecom_dag_new @ 2024-12-01 20:25:41.311394+00:00: manual__2024-12-01T20:25:41.311394+00:00, state:running, queued_at: 2024-1...

task_failure_DAG

ConditionsSeverity

Policy violates when ANY condition is met

Error

Configurations

Log query

resource.type="cloud_composer_environment" resource.labels.location="us-central1" resource.labels.environment_name="automl-composer" log_name="projects/axial-rigging-438817-h4/logs/airflow-worker" severity=ERROR

Notification rate limit

One notification per 5 minutes

Incident autoclose duration

1 day

Logs

SeverityDefault

Filter Search all fields and values

?

↺

↻

SEVERITY	TIMESTAMP	SUMMARY
> <div>Error</div>	2024-12-01 16:42:55.014 EST	airflow-worker Task failed with exception\nTraceback (most recent call last):\n File "/opt/python3.11/lib/python3.11/site-packages/airflow/models/taskinstance.py", ...
> <div>Error</div>	2024-12-01 16:42:55.084 EST	airflow-worker Failed to execute job 3708 for task schema_validation (Column CustomerCareCalls has values above the maximum allowed.; 168842)
> <div>Error</div>	2024-12-01 16:43:02.971 EST	airflow-worker Please set env variable SPARK_VERSION
> <div>Error</div>	2024-12-01 16:47:56.962 EST	airflow-worker Please set env variable SPARK_VERSION
> <div>Error</div>	2024-12-01 16:47:59.431 EST	airflow-worker Column CustomerCareCalls has values above the maximum allowed.
> <div>Error</div>	2024-12-01 16:47:59.432 EST	airflow-worker Validation Error: Column CustomerCareCalls has values above the maximum allowed.
> <div>Error</div>	2024-12-01 16:47:59.437 EST	airflow-worker Task failed with exception\nTraceback (most recent call last):\n File "/opt/python3.11/lib/python3.11/site-packages/airflow/models/taskinstance.py", ...

[ALERT - Error] task_failure_DAG for Cloud Composer Environment with {environment_name=automl-composer, location=us-central1, project_id=axial-rigging-438817-h4}

🔍 📄 📧

Inbox x

Google Cloud Alerting

VIEW INCIDENT Log alert fired Error Cloud Composer Environment with a log matching the query has appeared Start time Dec 1, 2024 at 9:42PM UTC (less than 1 sec

Sun, Dec 1, 4:42 PM (21 hours ago) ☆

Google Cloud Alerting <alerting-noreply@google.com>

to me

Sun, Dec 1, 4:48 PM (21 hours ago) ☆ 😊 ↶ ⋮

Google Cloud

VIEW INCIDENT

Log alert fired

Error

Cloud Composer Environment with a log matching the query has appeared

Start time

Dec 1, 2024 at 9:42PM UTC (~5 minutes ago)

Policy

task_failure_DAG

Project

axial-rigging-438817-h4

Condition

Log match condition

environment_name : automl-composer

location : us-central1

project_id : axial-rigging-438817-h4

2:27



50303



Yesterday • 4:42 PM

[ALERT - Error] task_failure_DAG for Cloud Composer Environment with {environment_name=automl-composer, location=us-central1, project_id=axial-rigging-438817-h4} ALRT-Error task_failure_DAG Log match condition

Unread

[ALERT - Error] task_failure_DAG for Cloud Composer Environment with {environment_name=automl-composer, location=us-central1, project_id=axial-rigging-438817-h4} ALRT-Error task_failure_DAG Log match condition

[ALERT - Error] failure_of_DAG for Cloud Composer Environment with {environment_name=automl-composer, location=us-central1, project_id=axial-rigging-438817-h4} ALRT-Error failure_of_DAG Log match condition

4:48 PM

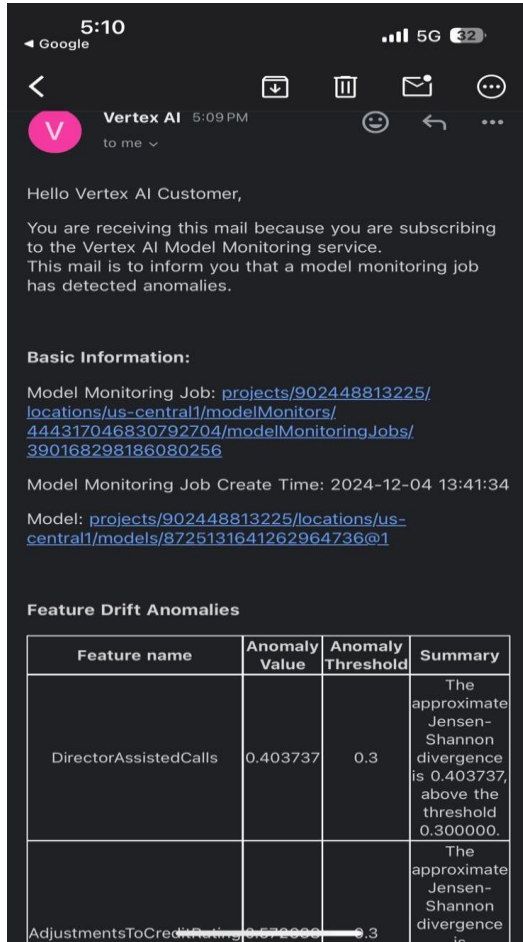


Text message



Model Monitoring:

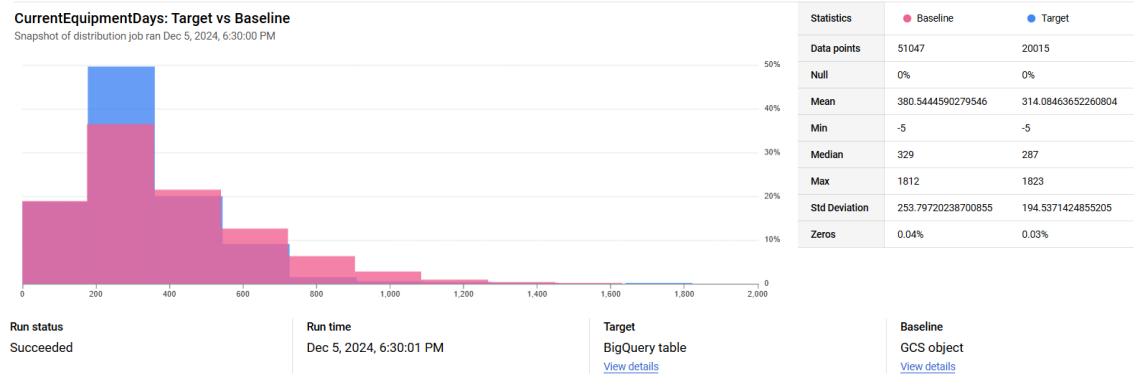
Model monitoring jobs set up and email alerts set up.



Drift detection jobs set up for analyzing feature drift, email alerts set up if drift score goes over the threshold.

Features	Baseline	Target	Alerts	Drift score	Threshold	Metric
AdjustmentsToCreditRating				0.58344	0.3	Jensen-Shannon
AgeHH1			—	0.00102	0.3	Jensen-Shannon
AgeHH2			—	0.00532	0.3	Jensen-Shannon
BuysViaMailOrder			—	0.02425	0.3	L-Infinity
CallWaitingCalls			—	0.03946	0.3	Jensen-Shannon
CreditRating				0.49963	0.3	L-Infinity
CurrentEquipmentDays			—	0.02856	0.3	Jensen-Shannon
CustomerCareCalls			—	0.19007	0.3	Jensen-Shannon
CustomerID			—	0.13037	0.3	Jensen-Shannon
DirectorAssistedCalls				0.3237	0.3	Jensen-Shannon
EmailID			—	0.16255	0.3	L-Infinity
HandsetModels			—	0.0295	0.3	Jensen-Shannon
HandsetPrice				0.61517	0.3	Jensen-Shannon
HandsetRefurbished			—	0.01126	0.3	L-Infinity
HandsetWebCapable			—	0.04381	0.3	L-Infinity
Handsets			—	0.05549	0.3	Jensen-Shannon

INPUT FEATURE DRIFT	OUTPUT PREDICTION DRIFT	FEATURE ATTRIBUTION DRIFT	CONFIGURATION
Run status Succeeded	Run time Dec 5, 2024, 6:30:01 PM	Target BigQuery table View details	Baseline GCS object View details



Target vs Baseline - Feature attribution score



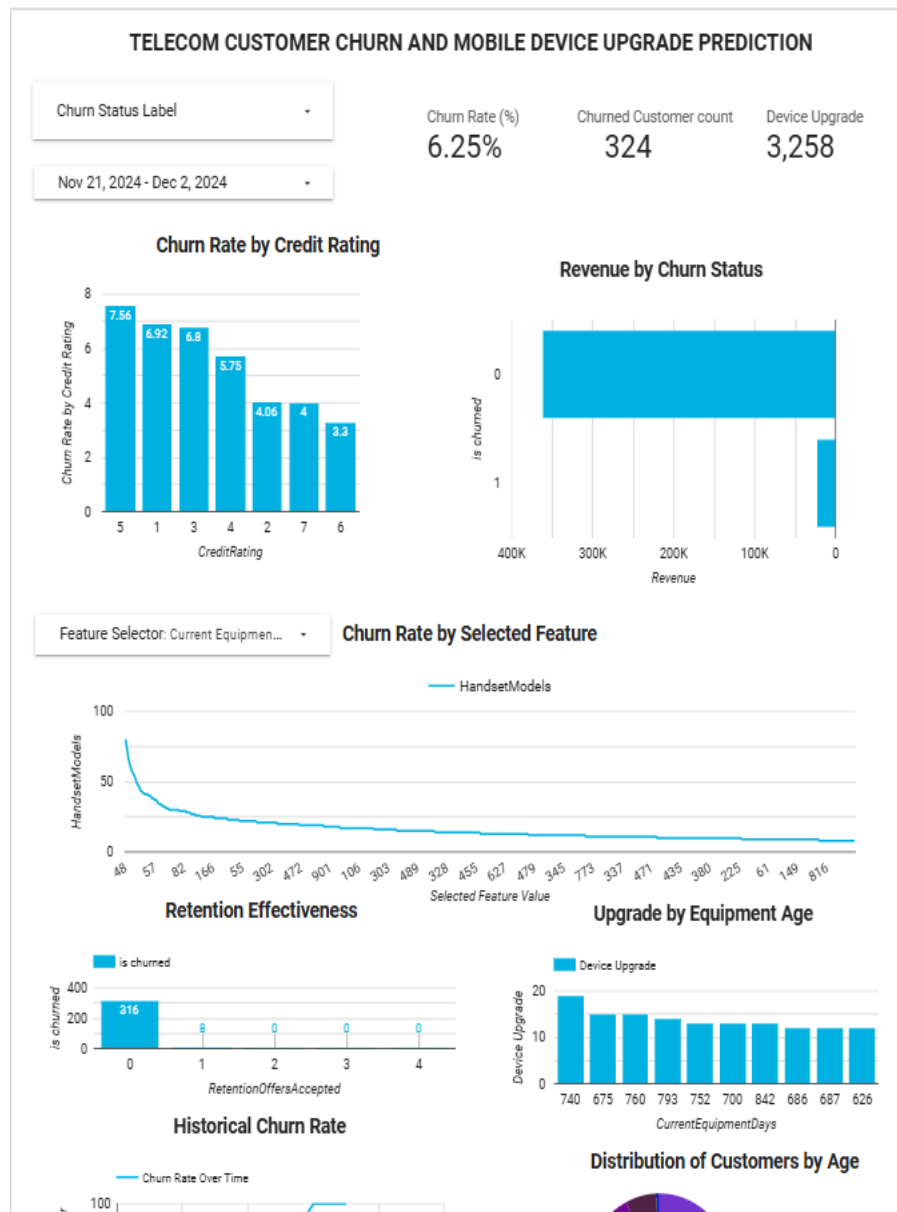
Retraining criteria:

Our architecture supports **on-demand retraining**, triggered by the addition of new data to the 'training data' bucket. The decision to include new data for retraining is informed by schema validation and anomaly detection tasks, which are part of the DAGs processing the incoming data. We also consider input feature drift and model performance (model drift) to prepare new training data and place it in the designated bucket. This action automatically triggers the training DAG, initiating the retraining process. By centralizing operations on GCP with storage buckets, Cloud Composer, and Vertex AI, our system maintains streamlined and automated workflows, eliminating the need for traditional CI/CD pipelines.

CI/CD:

We determined that integrating with Git for CI/CD workflows related to code changes is unnecessary for our project as it is fully deployed on Google Cloud Platform (GCP), leveraging Vertex AI's AutoML for model training and predictions. Since Vertex AI AutoML handles the modeling process, no custom model code is required. We have implemented automatic triggers based on incoming data, ensuring that Dataflow pipelines and directed acyclic graphs (DAGs) in Cloud Composer are executed whenever new data is added to the GCP storage buckets. The logic for these DAGs is predefined and fixed, automating the entire workflow without requiring code modifications.

Front End – Looker Analytics Dashboard



Since the telecom companies receive data in big batches, a front end is designed to display analytics of the batches rather than per customer user input-prediction design.