Write – Up of the final project – Survival rate of Ship Accidents

Initially, I selected the dataset of Titanic dataset from Kaggle.

https://www.kaggle.com/datasets/yasserh/titanic-dataset

It has 1089 instances and 13 features. The main goal of this project is to find whether a passenger will survive based on gender, age and class.

Implementation on GCP:

- I started by making a bucket and adding the dataset file to it.

- I also chose the tabular dataset type in the vertex AI dataset section.

- I trained the dataset with "Sensitive" as the target column,

- As well as then deployed it to the end point. Following its deployment into the endpoint.

- By using the project id and endpoint number as a reference, I wrote the pertinent code in Sage maker and forecasted the online instances. I saw that the prediction was accurate.

PKL Creation:

In the sage maker, I imported the basic libraries to read the dataset and performed certain basic statistics to understand the dataset.

And taken the 'Label', 'Brand', "Name", "Price", "Rank", "Ingredients", "Combination"as model features and "Sensitive" as model_target. And after that cleansed the dataset. And did the Count vectorizer where the data is converted into the binary form. Then encoded the dataset into the label.

I have split the dataset into 90% as test data and 10% as train data.

After that, the dataset is divided into class '1' and '0'. and concatenated it. After that shuffled the dataset.

And fitted the dataset into the pipeline. And calculated the Accuracy score. And generated the Model pkl file. After the model pkl file is created.

In the GCP, I have imported the model pkl file into the bucket. And in the model registry I have imported the pkl file and deployed it to the endpoint.

By using the endpoint number and project Id. I have tested the Predictions in the sage Maker by giving the online instances.

Things you found helpful and something you found unhelpful along the way

I took the help of google and learnt how to upload a dataset to bucket and train it. Labs helped me to work smoothly on the project. Using GCP is little difficult for me as it took more time for model creation and the coupons are completed to add the billing. However, I learnt how GCP works and how to train a model using it.

A summary of the lessons learned from working on the project

I got the chance to explore the features of vertex AI like bucket, dataset and about pkl file. Now I am good at training the model and making predictions using GCP

If you could do it again, is there anything that you would either do differently or change your project idea itself?

No, I am good with my current project as it involves using both gcp and jyupiter notebook

Google Cloud   final project ▼

Search (/) for resources, docs, products, and more   Q Search   7

Vertex AI

← project > Version 1 ▼   ⊞ VIEW DATASET   ⬇ EXPORT   🎓 LEARN

Datasets
Labeling tasks

MODEL DEVELOPMENT ∧

Training
Experiments
Metadata

DEPLOY AND USE ∧

Model Registry
Online prediction
Batch predictions
Matching Engine
Marketplace

⫷

EVALUATE   DEPLOY & TEST   BATCH PREDICT   VERSION DETAILS

← untitled_6340574916130848995 ▼   ⊸ COMPARE   ✚ CREATE EVALUATION

≡ Filter

All labels	0.892
0	0.897
1	0.853

Confidence threshold ❓ —————●———— 0.5

All labels

PR AUC ❓	0.892
ROC AUC ❓	0.891
Log loss ❓	0.418
F1 score ❓	0.8105263
Precision ❓	81.1%
Recall ❓	81.1%
Created	May 1, 2023, 10:41:01 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 some example scenarios ↗ to learn how evaluation metrics can be used.