

Powered by Plain Concepts

#STechDay2021

SINGULARITY TECH DAY_2021

The era of AI and Cognitive Services

Vertex AI

ORGANIZATION



SPONSORS



COLLABORATORS



Microsoft

Glovo⁹

cabify



PROSEGUR



MAHOU
SANMIGUEL

Google

sidra
Data Platform





Juliet Moreiro

Cloud Partner Engineer at Google



Machine Learning APIs:

Ready to Go



Cloud
Vision API



Cloud
Speech API



Cloud
Translation API



Cloud Natural
Language API



Cloud Video
Intelligence API

AutoML:

Bring Your Own Data
(We Do the Rest)



ML Frameworks:

Total Control



TensorFlow



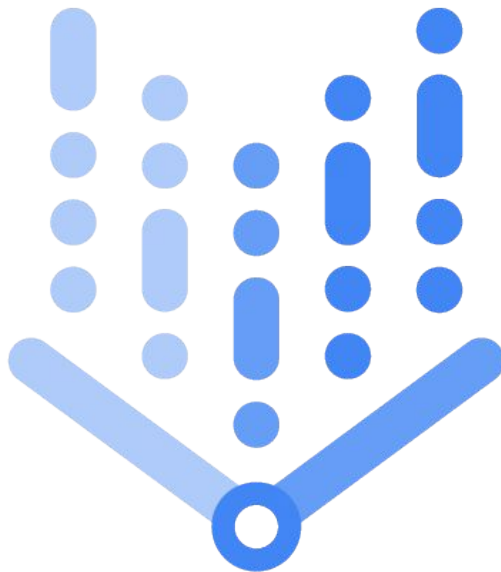
BigQuery ML



Spark ML



Cloud Dataproc



Vertex AI

Fun fact: The logo includes a message in Morse Code



#STechDay2021

Vertex AI

Applications

Vision and Video

Conversation

Language

Structured Data

Custom machine learning

Workbench

AutoML

Matching Engine

Prediction

ML Metadata

Data Labeling

Training

Explainable AI

Feature Store

Model Monitoring

Experiments

Vizier

Pipelines

Demo time!



Custom training

Pre-built containers

TensorFlow
scikit-learn
XGBoost

Custom containers

Push to Artifact
Registry or Docker
Hub

Configure

Training code runs
on one or more
VMs. Can use GPU.

Custom training job

Worker pool,
service account, etc

Managed Pipelines



Easy to use Python SDKs

Build your Pipelines using the battle-tested and easy-to-use KFP SDK and TFX SDK



Scalable

Run as many pipelines on as much data as you want without having to worry about compute resources.



Cost effective

Only pay for the pipelines you run and the resources they use.



Metadata Tracking and Lineage

Pipelines will automatically store metadata about every artifact produced by the Pipelines



Secure

Will be integrated with GCP security features like IAM, VPC-SC, and CMEK.

Feature Store

A rich feature repository to serve, share and re-use ML features.



Share and reuse ML features across use cases

Centralized feature repository with easy APIs to search & discover features, fetch them for training/serving and manage permissions.



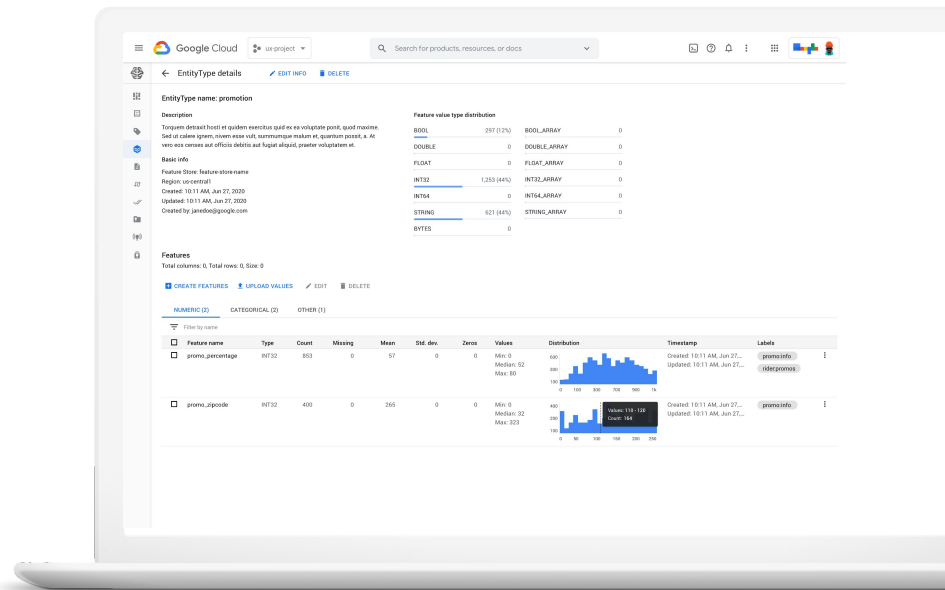
Serve ML Features at scale with low latency

Offload the operational overhead of handling infrastructure for low latency scalable feature serving.



Alleviate training serving skew

- Compute feature values once, re-use for training and serving
- Track & monitor for drift and other quality issues



Easy and proactive **monitoring** of model performance

For models deployed in the Vertex Prediction service



Monitor and alert

Monitor signals for model's predictive performance, and alert when those signals deviate.



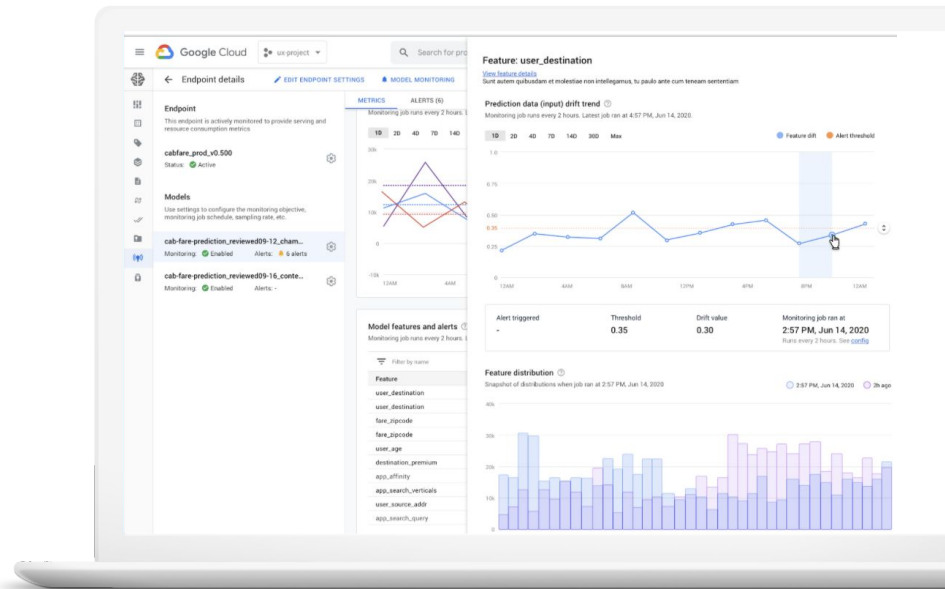
Diagnose

Help identify the cause for the deviation i.e. what changed, how and how much?



Update Model

Trigger model re-training pipeline or collect relevant training data to address performance degradation.



ML Metadata

#STechDay2021

Artifact, lineage, and execution tracking for your ML workflow



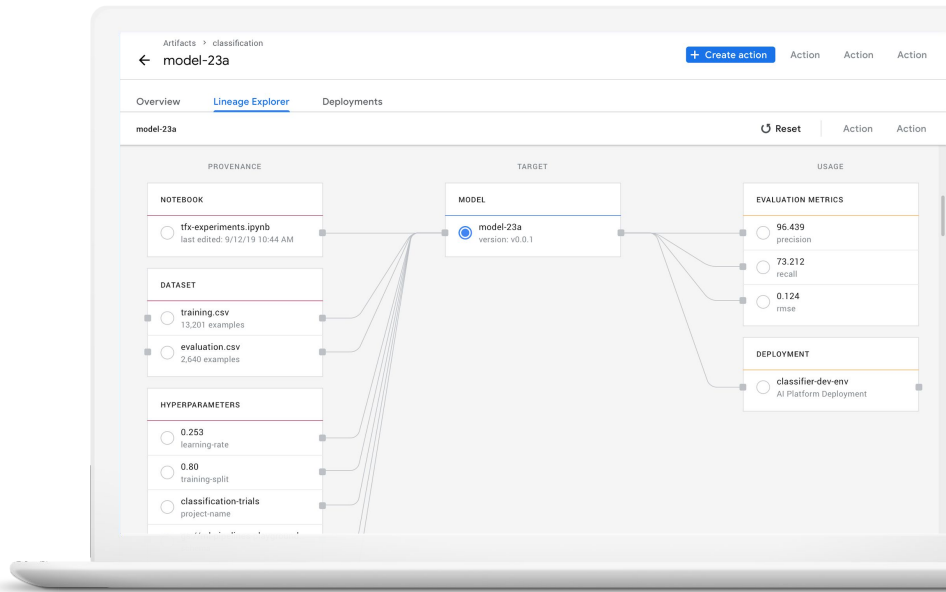
Automatically track inputs and outputs to all components in an ML pipeline, and their lineage.



Visualize the workflow for faster debugging with a DAG of all related executions.



Manage artifacts by projects, group by experiments, and track the usage of datasets and models in your organization



Matching Engine

#STechDay2021



Faster i.e low latency

Find nearest neighbors in a few millisecond



The most scalable

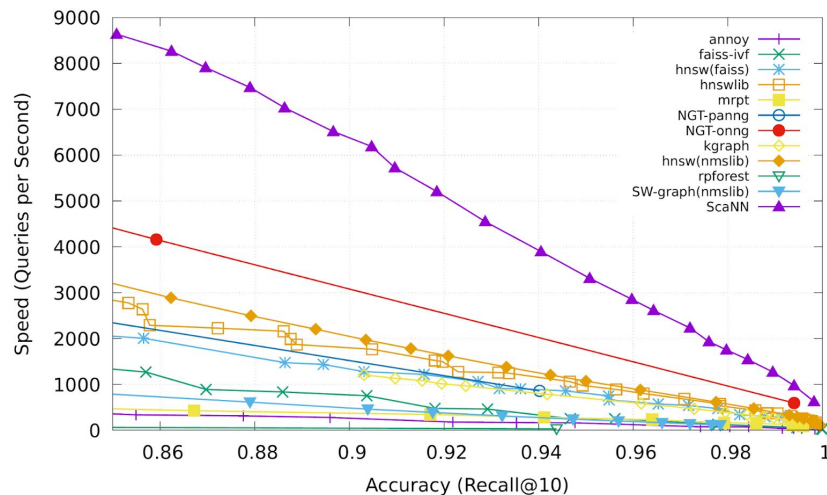
Scales to billions of vectors



Cheaper

Requires fewer VMs to serve the same workload

- 1/4th the CPU consumption of **faiss**
- 1/3rd the memory consumption of **nmslib**



Google's technology (labelled **ScaNN**) compared with popular ANN services



Automatic Hyperparameter tuning

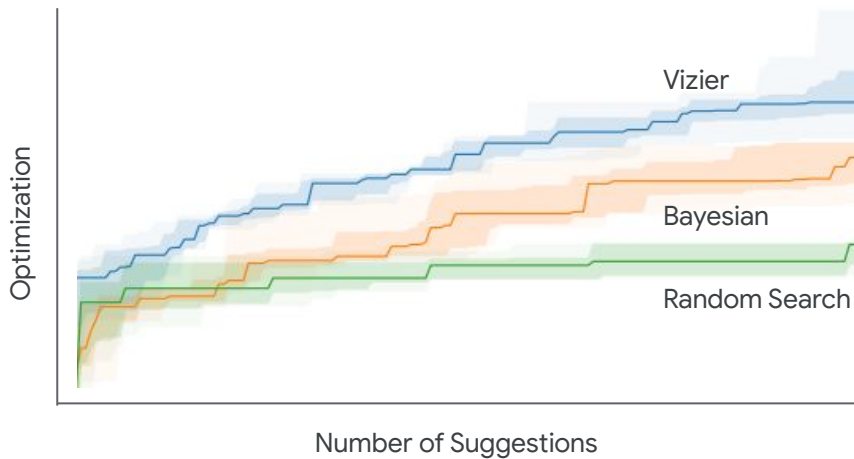
Easily sweep through different configurations to optimize your deep ML models using Vertex AI's HPTuning Jobs



Cloud Vizier Optimizer

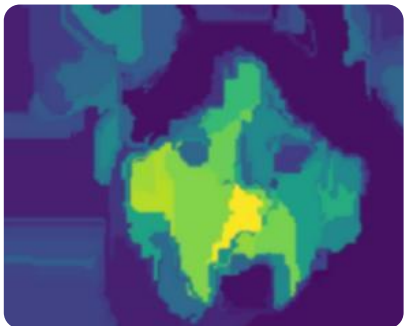
Google's generalized optimization engine, used by major Google apps for better and more efficient optimization optimization

Vizier optimizers better and faster



Explainable AI reveals the ‘why’ behind your model & predictions

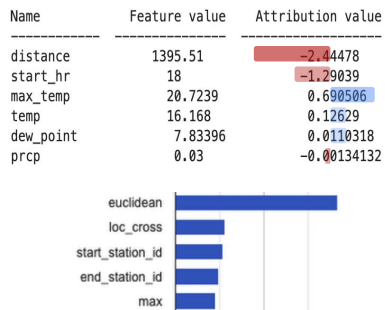
Images



Explanations tell you:

What **image pixels or regions** most contributed to the model's classification?

Tabular



How much did each **feature column** contribute to a single prediction or the model overall?

Text

The cake tastes delicious!

Sentiment score: 0.9

How much did each **word or token** contribute to the text classification?

Managed **experiment tracking** for faster model selection

#STechDay2021

Preview

Track, analyze, and discover ML experiments within your organization.

Customers can..



Track all runs and metrics in a central place

A central dashboard for training & evaluation runs across frameworks and environments in a single-pane view.



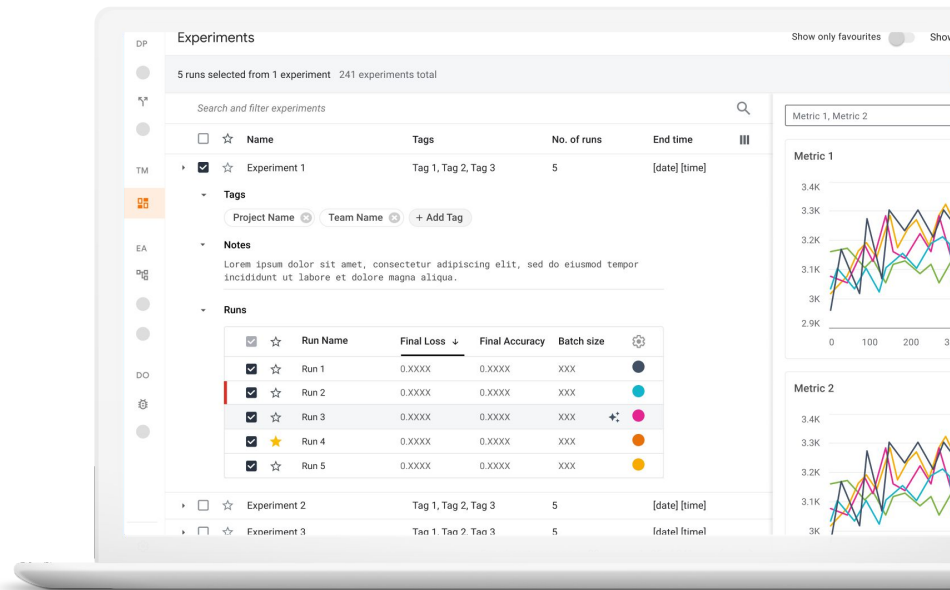
Visualize and compare multiple experiments

Analyze different training runs with rich, built-in visualizations. Quickly compare experiments and select the best model.



Share experiments and collaborate effectively

Search and discover experiments across the organization with ease; share findings easily.



BigQuery ML

#STechDay2021

① FEATURES & INFO 📄 SHORTCUT 🗑️ DISABLE EDITOR TABS

Explorer + ADD DATA

🔍 Type to search ?

Viewing pinned projects.

- ▼ jmoreirodemos
 - 🔍 Saved queries (3)
 - Create Model**
 - Model Predict
 - Model Predict Subset
- ▶ BQML
- ▶ HappyDataset
- ▶ telco
- ▶ bigquery-public-data

EDITOR ✕ CREATE ... ✕

▶ RUN 📄 SAVE 👤 SHARE ⌚ SCHEDULE ⚙️ MORE

```
1 CREATE MODEL BQML.model_1 OPTIONS(  
2   model_type = 'kmeans',  
3   num_clusters = 5,  
4   distance_type = 'euclidean'  
5 ) AS  
6 SELECT  
7   Annual_Revenue,  
8   timeOnScreen,  
9   UniqueScreenViews,  
10  Loyalty_Program,  
11  Lifetime_Value,  
12  Age  
13 FROM  
14  `caramel-howl-182418.CustomerSegmentation_Clustering.Dataset`;
```

SINGULARITY
TECH DAY_2021

Thank you!

