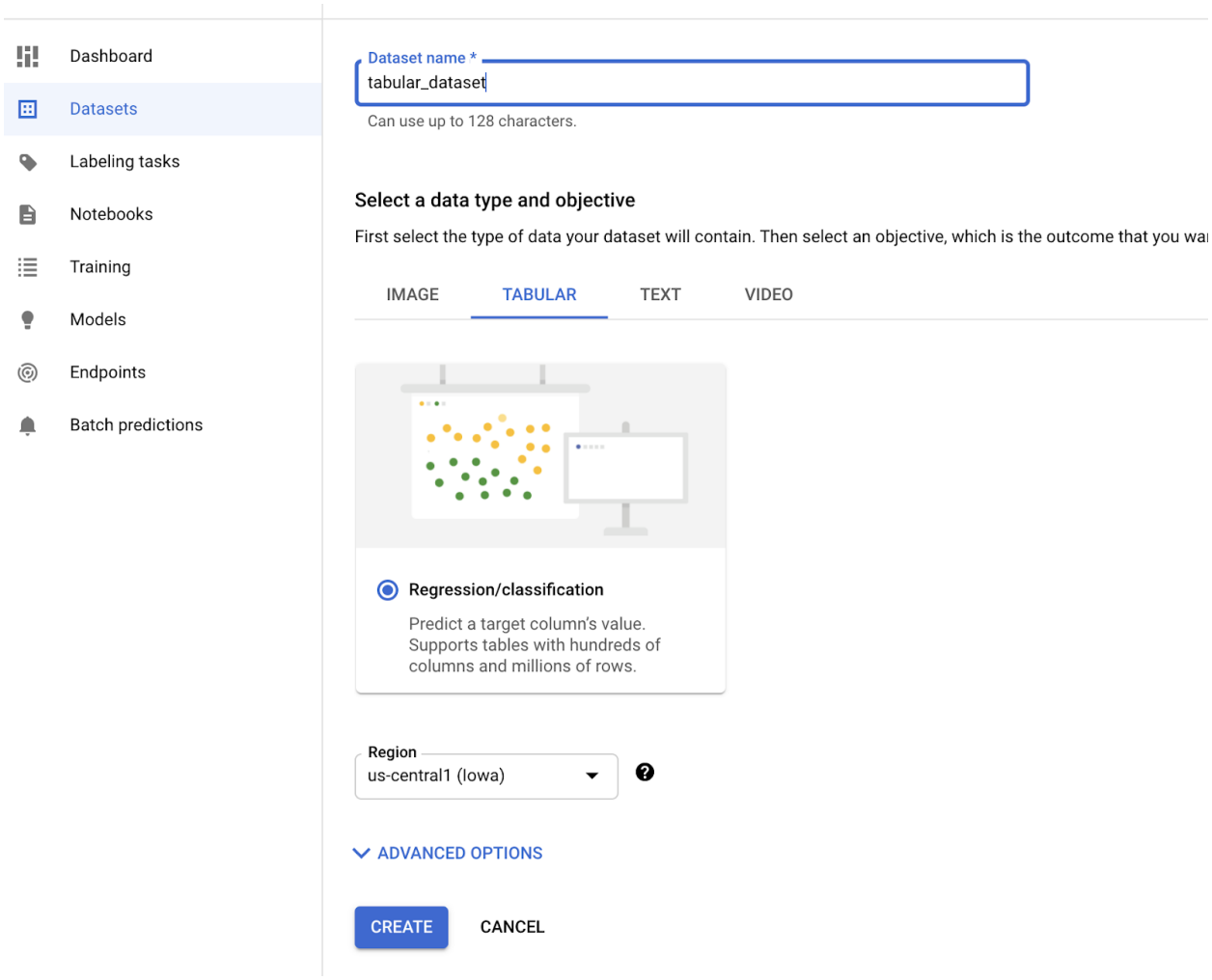
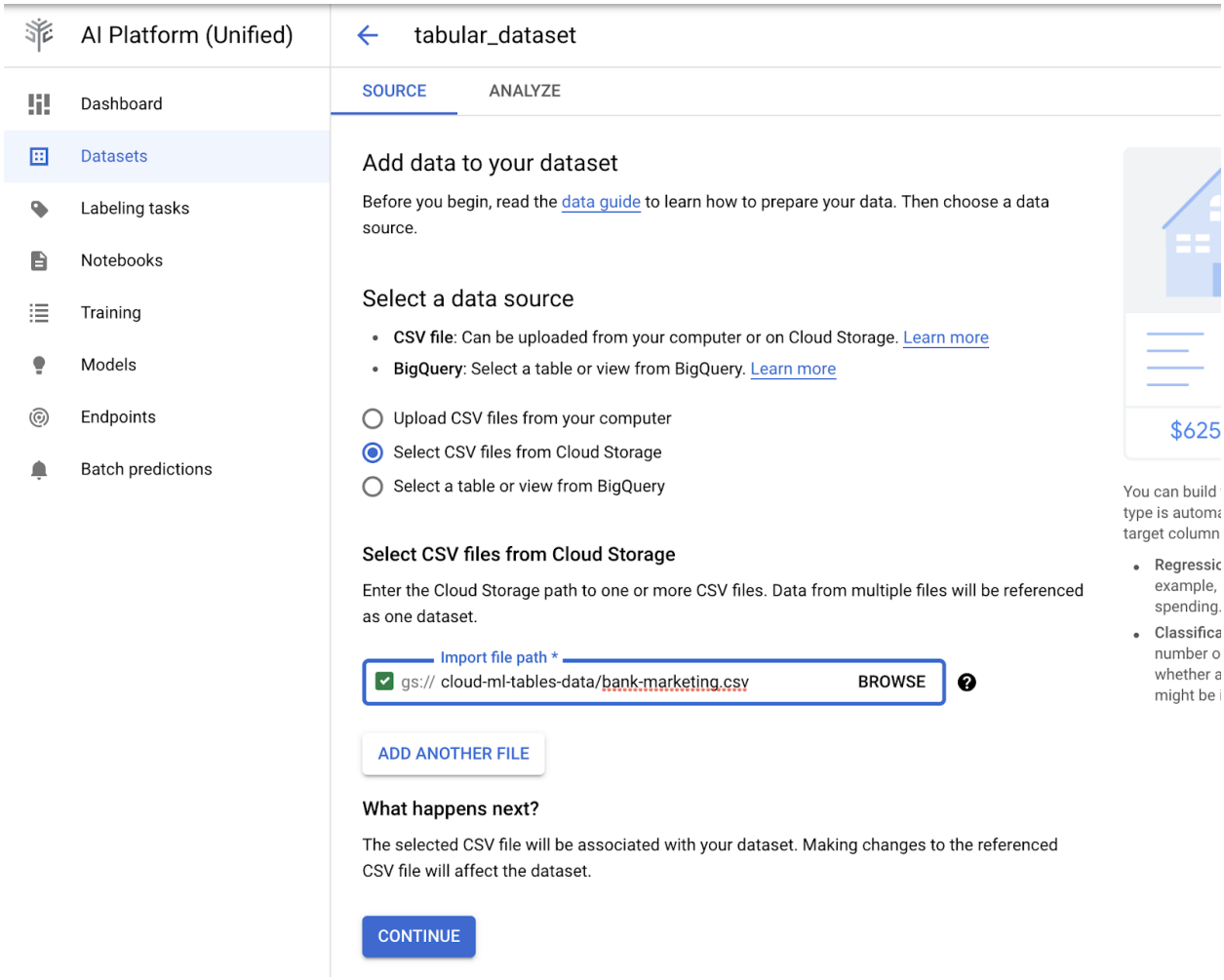
Tabular Dataset




AI Platform (Unified)
Dashboard
Datasets
Labeling tasks
Notebooks
Training
Models
Endpoints
Batch predictions
tabular_dataset
SOURCE
ANALYZE
Add data to your dataset
Before you begin, read the data guide to learn how to prepare your data. Then choose a data source.
Select a data source
CSV file: Can be uploaded from your computer or on Cloud Storage. Learn more
BigQuery: Select a table or view from BigQuery. Learn more
Upload CSV files from your computer
Select CSV files from Cloud Storage
Select a table or view from BigQuery
Select CSV files from Cloud Storage
Enter the Cloud Storage path to one or more CSV files. Data from multiple files will be referenced as one dataset.
Import file path *
gs:// cloud-ml-tables-data/bank-marketing.csv
BROWSE
ADD ANOTHER FILE
What happens next?
The selected CSV file will be associated with your dataset. Making changes to the referenced CSV file will affect the dataset.
CONTINUE
$625
You can build
type is automa
target column
Regressic
example,
spending.
Classifica
number o
whether a
might be

tabular_dataset

SOURCE | ANALYZE

Dataset Info

Created: Mar 07, 2021 4:52 PM

Dataset format: CSV

Dataset location: gs://cloud-ml-.../bank-marketing.csv

Summary

Total columns: 17

Total rows: 45,211

General statistics generated by Mar 07, 2021 5:00 PM GENERATE STATISTICS

Enter property name or value

Field Name	Missing % (count)	Distinct values
Age	-	77
Balance	-	7168
Campaign	-	48
Contact	-	3
Day	-	31
Default	-	2
Deposit	-	2
Duration	-	1573
Education	-	4
Housing	-	2
Job	-	12
Loan	-	2
MaritalStatus	-	3
Month	-	12

Train new model

1. Choose training method
2. Define your model
3. Choose training options
4. Compute and pricing

START TRAINING CANCEL

Model name *
tabular_dataset_2021381522

Target column *
Deposit

Export test dataset to BigQuery

ADVANCED OPTIONS

CONTINUE

Train new model

- Choose training method
- Define your model
- Choose training options
- 4 Compute and pricing

START TRAINING CANCEL

Dataset: tabular_dataset

Objective *: Classification

Please refer to the pricing guide for more details (and available deployment options) for each method.

(•) AutoML
Train high-quality models with minimal effort and machine learning expertise. Just specify how long you want to train. Learn more

( ) Custom training (advanced)
Run your TensorFlow, scikit-learn, and XGBoost training applications in the cloud. Train with one of Google Cloud's pre-built containers or use your own. Learn more

CONTINUE

tabular_dataset_2021381522 VIEW DATASET EXPORT

EVALUATE DEPLOY & TEST BATCH PREDICTIONS MODEL PROPERTIES

Filter labels

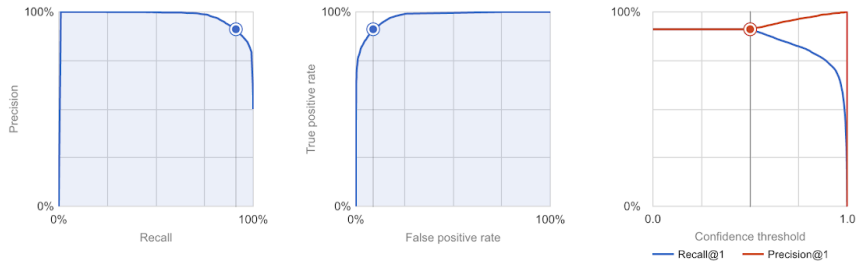
Confidence threshold 0.5

All labels	0
1	0.99154
2	0.61459

All labels

PR AUC	0.978
ROC AUC	0.976
Log loss	0.192
F1 score	0.91090214
Precision	91.1%
Recall	91.1%
Created	Mar 7, 2021, 6:20:53 PM

Use the slider to see which confidence threshold works best for your model on the precision-recall tradeoff curve. Learn more about these metrics and graphs



All labels	0
1	0.99154
2	0.61459

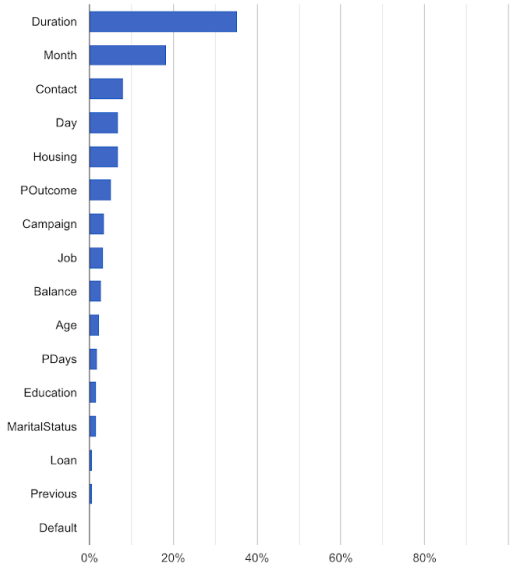
Confusion matrix

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). **Note that this table is limited to the 1 labels.** You can download the entire confusion matrix as a CSV file.

True label \ Predicted label	1	2
1	96%	4%
2	51%	49%

All labels	0
1	0.99154
2	0.61459

Feature Importance



Deploy to endpoint

1. Define your endpoint
2. Endpoint details

DEPLOY CANCEL

Create new endpoint Add to existing endpoint

Endpoint name *
tabular_data_endpoint

Model settings

tabular_dataset_2021381522

Traffic split *
100 %

Compute resources

Choose how compute resources will serve prediction traffic to your model

- **Autoscaling**: If you set a minimum and maximum, compute nodes will scale to meet traffic demand within those boundaries
- **No scaling**: If you only set a minimum, then that number of compute nodes will always run regardless of traffic demand (the maximum will be set to minimum)

Once scaling settings are set, they can't be changed unless you redeploy the model. Pricing guide

Minimum number of compute nodes *
1

Default is 1. If set to 1 or more, then compute resources will continuously run even without traffic demand. This can increase cost but avoid dropped requests due to node initialization.

Maximum number of compute nodes (optional)

Enter a number equal to or greater than the minimum nodes. Can reduce costs but may cause reliability issues for high traffic.

Machine type *
n1-standard-8, 8 vCPUs, 30 GiB memory