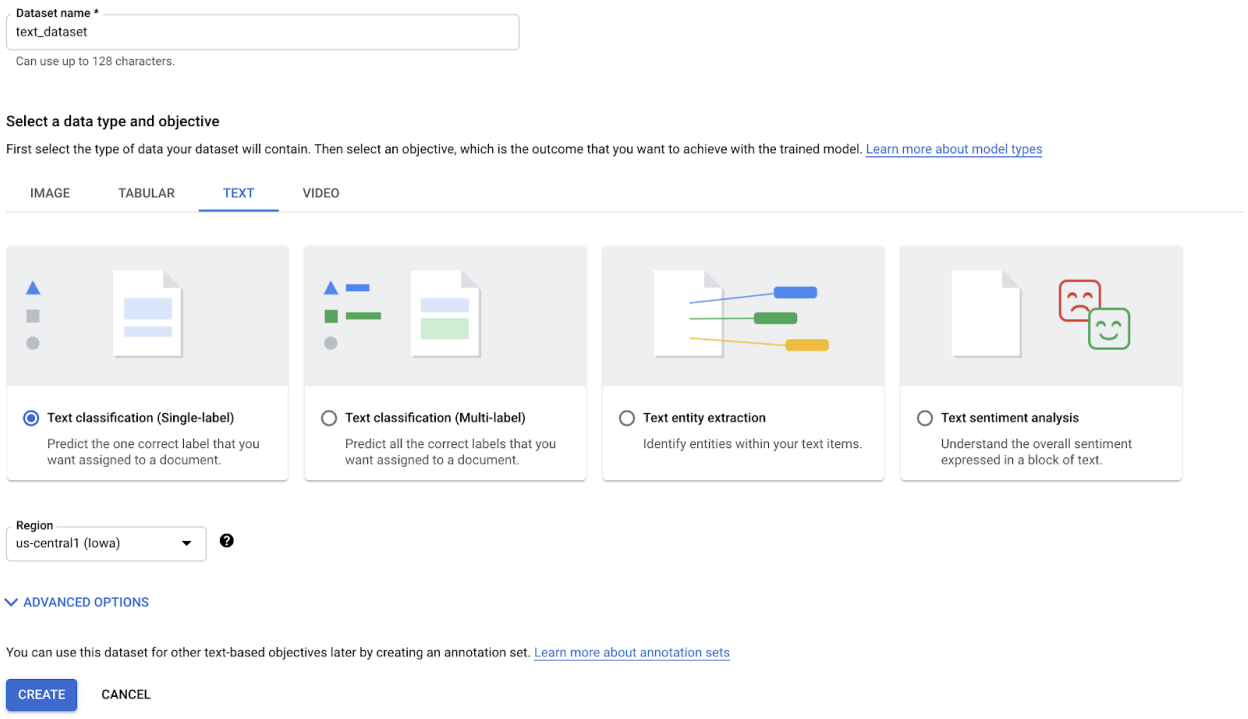
Text Dataset

```
teepika_ramasamymarimuthu@cloudshell:~ (direct-hope-306504)$ gsutil mb -p ${PROJECT_ID} -l us-central1 gs://${BUCKET}/
Creating gs://direct-hope-306504-lcm/...
teepika_ramasamymarimuthu@cloudshell:~ (direct-hope-306504)$
teepika_ramasamymarimuthu@cloudshell:~ (direct-hope-306504)$ gsutil -m cp -R gs://cloud-ml-data/NL-classification/happiness.csv gs://${BUCKET}/text/
Copying gs://cloud-ml-data/NL-classification/happiness.csv [Content-Type=text/csv]...
/ [1/1 files][  1.3 MiB/  1.3 MiB] 100% Done
Operation completed over 1 objects/1.3 MiB.
```



text_dataset text_dataset_tcn

IMPORT BROWSE ANALYZE

Add text documents to your dataset

Before you begin, read the data guide to learn how to prepare your data. Then choose an import method.

Select an import method

- **Upload text documents**: Recommended if you don't have labels yet
- **Import files**: Recommended if you already have labels. An import file is a list of Cloud Storage URIs to your text documents and optional data, like labels. Learn how to create an import file

- ○ Upload text documents from your computer
- ○ Upload import files from your computer
- ◉ Select import files from Cloud Storage

Select import files from Cloud Storage

Text documents referenced in the import files will be preprocessed and stored in a new Cloud Storage bucket (charges apply)

Import file path * gs:// direct-hope-306504-lcm/text/happiness.c BROWSE

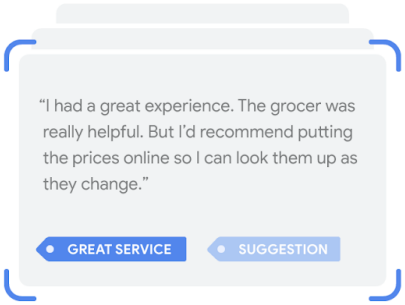
Data split Automatic

ADD ANOTHER FILE

What happens next?

You'll be emailed after the text documents are imported and your dataset is ready

CONTINUE



Text classification models predict one or more labels for a document.

If you want to get sentiment analysis, entity analysis, content classification, and syntax analysis with Google's pre-trained models, try the **Natural Language API**. Learn more

Train new model

1. Choose training method
2. Define your model

START TRAINING CANCEL

Dataset
text_dataset

Annotation set
text_dataset_tcn

Objective
Text classification (Single-label)

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. **AutoML training automatically ends when you model stops improving.** 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

Train new model

1. ✓ Choose training method
2. Define your model

START TRAINING CANCEL

Model name *
text_dataset_20213543628

Data split

(•) Randomly assigned ( ) Manual (Advanced)

Your dataset will be automatically randomized and split into training, validation, and test sets using the following ratios.

Training	Validation	Test
80 %	10 %	10 %

Encryption

[ ] Use a customer-managed encryption key (CMEK)

SHOW LESS

text_dataset_20213543628 VIEW DATASET

EVALUATE | DEPLOY & TEST | BATCH PREDICTIONS | MODEL PROPERTIES

Filter labels

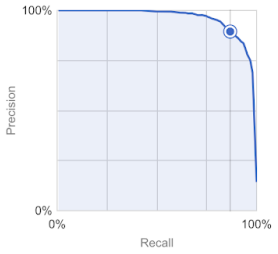
All labels	0
affection	0.99186
bonding	0.97597
achievement	0.96046
nature	0.92880
exercise	0.87311
enjoy_the_moment	0.85618
leisure	0.85253

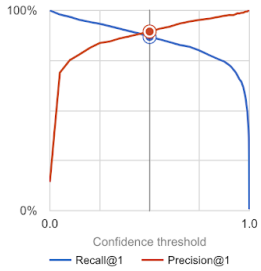
Confidence threshold 0.5

All labels

Average precision	0.962
Precision	89.5%
Recall	86.8%
Created	Mar 5, 2021, 12:51:00 AM
Total items	11,947
Training items	9,555
Validation items	1,207
Test items	1,185

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





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 th labels.** You can download the entire confusion matrix as a CSV file.

True label \ Predicted label	enjoy_the_moment	leisure	achievement	affection	bonding	nature	exercise
enjoy_the_moment	72%	4%	17%	4%	1%	–	1%
leisure	12%	71%	13%	1%	1%	–	1%

Test your model PREVIEW

I'm going for a holiday trip to San Francisco

PREDICT

Filter labels

affection	0.020
achievement	0.080
enjoy_the_moment	0.788
bonding	0.002
leisure	0.109
nature	0.000
exercise	0.000