```
!pip install transformers
!pip install datasets
!pip install evaluate
```

```
Requirement already satisfied: transformers in /usr/local/lib/python3.11/dist-packages (4.50.3)
Requirement already satisfied: filelock in /usr/local/lib/python3.11/dist-packages (from transformers) (3.18.0)
Requirement already satisfied: huggingface-hub<1.0,>=0.26.0 in /usr/local/lib/python3.11/dist-packages (from transformers) (0.30.1)
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.11/dist-packages (from transformers) (2.0.2)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.11/dist-packages (from transformers) (24.2)
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.11/dist-packages (from transformers) (6.0.2)
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.11/dist-packages (from transformers) (2024.11.6)
Requirement already satisfied: requests in /usr/local/lib/python3.11/dist-packages (from transformers) (2.32.3)
Requirement already satisfied: tokenizers<0.22,>=0.21 in /usr/local/lib/python3.11/dist-packages (from transformers) (0.21.1)
Requirement already satisfied: safetensors>=0.4.3 in /usr/local/lib/python3.11/dist-packages (from transformers) (0.5.3)
Requirement already satisfied: tqdm>=4.27 in /usr/local/lib/python3.11/dist-packages (from transformers) (4.67.1)
Requirement already satisfied: fsspec>=2023.5.0 in /usr/local/lib/python3.11/dist-packages (from huggingface-hub<1.0,>=0.26.0->transf
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.11/dist-packages (from huggingface-hub<1.0,>=0.26
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.11/dist-packages (from requests->transformers) (3.4
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-packages (from requests->transformers) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-packages (from requests->transformers) (2.3.0)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.11/dist-packages (from requests->transformers) (2025.1.31
Collecting datasets
  Downloading datasets-3.5.0-py3-none-any.whl.metadata (19 kB)
Requirement already satisfied: filelock in /usr/local/lib/python3.11/dist-packages (from datasets) (3.18.0)
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.11/dist-packages (from datasets) (2.0.2)
Requirement already satisfied: pyarrow>=15.0.0 in /usr/local/lib/python3.11/dist-packages (from datasets) (18.1.0)
Collecting dill<0.3.9,>=0.3.0 (from datasets)
  Downloading dill-0.3.8-py3-none-any.whl.metadata (10 kB)
Requirement already satisfied: pandas in /usr/local/lib/python3.11/dist-packages (from datasets) (2.2.2)
Requirement already satisfied: requests>=2.32.2 in /usr/local/lib/python3.11/dist-packages (from datasets) (2.32.3)
Requirement already satisfied: tqdm>=4.66.3 in /usr/local/lib/python3.11/dist-packages (from datasets) (4.67.1)
Collecting xxhash (from datasets)
  Downloading xxhash-3.5.0-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (12 kB)
Collecting multiprocess<0.70.17 (from datasets)
  Downloading multiprocess-0.70.16-py311-none-any.whl.metadata (7.2 kB)
Collecting fsspec<=2024.12.0,>=2023.1.0 (from fsspec[http]<=2024.12.0,>=2023.1.0->datasets)
  Downloading fsspec-2024.12.0-py3-none-any.whl.metadata (11 kB)
Requirement already satisfied: aiohttp in /usr/local/lib/python3.11/dist-packages (from datasets) (3.11.15)
Requirement already satisfied: huggingface-hub>=0.24.0 in /usr/local/lib/python3.11/dist-packages (from datasets) (0.30.1)
Requirement already satisfied: packaging in /usr/local/lib/python3.11/dist-packages (from datasets) (24.2)
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.11/dist-packages (from datasets) (6.0.2)
Requirement already satisfied: aiohappyeyeballs>=2.3.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp->datasets) (2.6.1)
Requirement already satisfied: aiosignal>=1.1.2 in /usr/local/lib/python3.11/dist-packages (from aiohttp->datasets) (1.3.2)
Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp->datasets) (25.3.0)
Requirement already satisfied: frozenlist>=1.1.1 in /usr/local/lib/python3.11/dist-packages (from aiohttp->datasets) (1.5.0)
Requirement already satisfied: multidict<7.0,>=4.5 in /usr/local/lib/python3.11/dist-packages (from aiohttp->datasets) (6.3.1)
Requirement already satisfied: propcache>=0.2.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp->datasets) (0.3.1)
Requirement already satisfied: yarl<2.0,>=1.17.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp->datasets) (1.18.3)
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.11/dist-packages (from huggingface-hub>=0.24.0->d
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.11/dist-packages (from requests>=2.32.2->datasets)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-packages (from requests>=2.32.2->datasets) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-packages (from requests>=2.32.2->datasets) (2.3.0
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.11/dist-packages (from requests>=2.32.2->datasets) (2025.
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.11/dist-packages (from pandas->datasets) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.11/dist-packages (from pandas->datasets) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.11/dist-packages (from pandas->datasets) (2025.2)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.8.2->pandas->datasets) (1
Downloading datasets-3.5.0-py3-none-any.whl (491 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 491.2/491.2 kB 35.4 MB/s eta 0:00:00
Downloading dill-0.3.8-py3-none-any.whl (116 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 116.3/116.3 kB 13.0 MB/s eta 0:00:00
```

```
!jupyter nbconvert --ClearOutputPreprocessor.enabled=True --inplace LORA.ipynb
```

```
from datasets import load_dataset

dataset = load_dataset("tweet_eval", "offensive")
```

```
README.md: 100% 23.9k/23.9k [00:00<00:00, 1.92MB/s]
train-00000-of-00001.parquet: 100% 1.02M/1.02M [00:00<00:00, 25.9MB/s]
test-00000-of-00001.parquet: 100% 93.7k/93.7k [00:00<00:00, 8.53MB/s]
validation-00000-of-00001.parquet: 100% 122k/122k [00:00<00:00, 7.55MB/s]
Generating train split: 100% 11916/11916 [00:00<00:00, 198933.78 examples/s]
Generating test split: 100% 860/860 [00:00<00:00, 42970.84 examples/s]
Generating validation split: 100% 1324/1324 [00:00<00:00, 57086.48 examples/s]
```

```
dataset
```

```
DatasetDict({
    train: Dataset({
        features: ['text', 'label'],
        num_rows: 11916
    })
    test: Dataset({
        features: ['text', 'label'],
        num_rows: 860
    })
    validation: Dataset({
        features: ['text', 'label'],
        num_rows: 1324
    })
})
```

```
id2label = {0: "NotOffensive", 1: "Offensive"}
label2id = {"NotOffensive":0, "offensive":1}
```

```
from transformers import AutoModelForSequenceClassification, AutoTokenizer
```

```
model_checkpoint = 'distilbert-base-uncased'
```

```
model = AutoModelForSequenceClassification.from_pretrained(
    model_checkpoint,
    num_labels = 2,
    id2label = id2label,
    label2id = label2id
)
```

```
config.json: 100% 483/483 [00:00<00:00, 51.8kB/s]
Xet Storage is enabled for this repo, but the 'hf_xet' package is not installed. Falling back to regular HTTP download. For better perfo
WARNING:huggingface_hub.file_download:Xet Storage is enabled for this repo, but the 'hf_xet' package is not installed. Falling back to r
model.safetensors: 100% 268M/268M [00:00<00:00, 315MB/s]
Some weights of DistilBertForSequenceClassification were not initialized from the model checkpoint at distilbert-base-uncased and are ne
You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.
```

```
tokenizer = AutoTokenizer.from_pretrained("distilbert-base-uncased")
```

```
tokenizer_config.json: 100% 48.0/48.0 [00:00<00:00, 1.82kB/s]
vocab.txt: 100% 232k/232k [00:00<00:00, 5.59MB/s]
tokenizer.json: 100% 466k/466k [00:00<00:00, 15.3MB/s]
```

```
def tokenize_function(data):
  return tokenizer(data["text"], padding = "max_length", truncation = True)

Tokenized_data = dataset.map(tokenize_function,batched = True)
```

```
Map: 100% 11916/11916 [00:05<00:00, 2101.25 examples/s]
Map: 100% 860/860 [00:00<00:00, 1669.84 examples/s]
Map: 100% 1324/1324 [00:00<00:00, 2914.04 examples/s]
```

```
import evaluate

accuracy = evaluate.load("accuracy")
```

```
Downloading builder script: 100%  4.20k/4.20k [00:00<00:00, 304kB/s]
```

```
import numpy as np


def compute_metrics(p):
    predictions = np.argmax(p.predictions, axis=1)
    return {"accuracy": accuracy.compute(predictions=predictions, references=p.label_ids)}



from peft import PeftModel, PeftConfig, get_peft_model, LoraConfig


peft = LoraConfig(
    task_type= "SEQ_CLS",
    r=4,
    lora_alpha = 32,
    lora_dropout = 0.01,
    target_modules = ['q_lin']
)


peft
```

```
LoraConfig(task_type='SEQ_CLS', peft_type=<PeftType.LORA: 'LORA'>, auto_mapping=None, base_model_name_or_path=None, revision=None,
 inference_mode=False, r=4, target_modules={'q_lin'}, exclude_modules=None, lora_alpha=32, lora_dropout=0.01, fan_in_fan_out=False,
 bias='none', use_rslora=False, modules_to_save=None, init_lora_weights=True, layers_to_transform=None, layers_pattern=None,
 rank_pattern={}, alpha_pattern={}, megatron_config=None, megatron_core='megatron.core', loftq_config={}, eva_config=None,
 use_dora=False, layer_replication=None, runtime_config=LoraRuntimeConfig(ephemeral_gpu_offload=False), lora_bias=False)
```

```
model = get_peft_model(model, peft)


model.print_trainable_parameters()
```

```
trainable params: 628,994 || all params: 67,584,004 || trainable%: 0.9307
```

```
from transformers import TrainingArguments, Trainer


training_args = TrainingArguments(
    output_dir = "/content/logs",
    learning_rate = 1e-3,
    per_device_train_batch_size=4,
    per_device_eval_batch_size = 8,
    num_train_epochs = 5,
    weight_decay = 0.01,
    evaluation_strategy = "epoch",
    save_strategy = "epoch",
    load_best_model_at_end = True,

    report_to="none"
)
```

```
/usr/local/lib/python3.11/dist-packages/transformers/training_args.py:1611: FutureWarning: `evaluation_strategy` is deprecated and will
  warnings.warn(
```

```
trainer = Trainer(
    model = model,
    args = training_args,
    train_dataset = Tokenized_data['train'],
    eval_dataset = Tokenized_data['validation'],
    tokenizer = tokenizer,
    compute_metrics = compute_metrics,
)
```

```
ion 5.0.0 for `Trainer.__init__`. Use `processing_class` instead.

ase models input arguments, if label_names is not given, label_names can't be set automatically within `Trainer`. Note that empty label_r
```

```
import os
os.environ["WANDB_DISABLED"] = "true"
```

```
trainer.train()
```

[14895/14895 37:31, Epoch 5/5]

Epoch	Training Loss	Validation Loss	Accuracy
1	0.610500	0.549541	{'accuracy': 0.7515105740181269}
2	0.584300	0.557035	{'accuracy': 0.7673716012084593}
3	0.561500	0.568429	{'accuracy': 0.7749244712990937}
4	0.535100	0.572076	{'accuracy': 0.7817220543806647}
5	0.482000	0.608261	{'accuracy': 0.7787009063444109}

```
TrainOutput(global_step=14895, training_loss=0.5602428373533354, metrics={'train_runtime': 2251.5057, 'train_samples_per_second':
```

```
import torch
```

```
test_tweets = [
    "You're such a loser, no one likes you.",
    "I really enjoyed the concert last night!",
    "Shut up already. No one cares what you think.",
    "Happy birthday! Wishing you all the best 🎉",
    "This is the dumbest thing I've ever seen.",
    "Grateful for everyone who supported me!",
    "You look disgusting. Stay home.",
    "The sunset today was absolutely beautiful!",
    "Go cry somewhere else, weakling.",
    "Just finished a great workout. Feeling amazing!"
]
```

```
for text in test_tweets:

    inputs = tokenizer.encode(text, return_tensors="pt").to(model.device)
    logits = model(inputs).logits

    predictions = torch.argmax(logits, dim=1)

    print(text + " - " + id2label[predictions.item()])
```

```
----------------------------
You're such a loser, no one likes you. - Offensive
I really enjoyed the concert last night! - NotOffensive
Shut up already. No one cares what you think. - NotOffensive
Happy birthday! Wishing you all the best 🎉 - NotOffensive
This is the dumbest thing I've ever seen. - Offensive
Grateful for everyone who supported me! - NotOffensive
You look disgusting. Stay home. - Offensive
The sunset today was absolutely beautiful! - NotOffensive
Go cry somewhere else, weakling. - NotOffensive
Just finished a great workout. Feeling amazing! - NotOffensive
```

```
test_dataset = Tokenized_data["test"]
test_results = trainer.evaluate(test_dataset)
print(test_results)
```

```
pred = trainer.predict(test_dataset)
print(pred.metrics)
```

```
{'eval_loss': 0.4943658113479614, 'eval_accuracy': {'accuracy': 0.7790697674418605}, 'eval_runtime': 13.2388, 'eval_samples_per_second':
{'test_loss': 0.4943658113479614, 'test_accuracy': {'accuracy': 0.7790697674418605}, 'test_runtime': 13.108, 'test_samples_per_second':
```

```
Start coding or generate with AI.
```