```
!pip install transformers==4.36.2 datasets scikit-learn pandas
```

```
Collecting transformers==4.36.2
  Downloading transformers-4.36.2-py3-none-any.whl.metadata (126 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 126.8/126.8 kB 2.5 MB/s eta 0:00:00
Collecting datasets
  Downloading datasets-3.5.1-py3-none-any.whl.metadata (19 kB)
Requirement already satisfied: scikit-learn in /usr/local/lib/python3.11/dist-packages (1.6.1)
Requirement already satisfied: pandas in /usr/local/lib/python3.11/dist-packages (2.2.2)
Requirement already satisfied: filelock in /usr/local/lib/python3.11/dist-packages (from transformers==4.36.2) (3.18.0)
Requirement already satisfied: huggingface-hub<1.0,>=0.19.3 in /usr/local/lib/python3.11/dist-packages (from transformers==4.36.2) (0
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.11/dist-packages (from transformers==4.36.2) (2.0.2)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.11/dist-packages (from transformers==4.36.2) (24.2)
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.11/dist-packages (from transformers==4.36.2) (6.0.2)
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.11/dist-packages (from transformers==4.36.2) (2024.11.6)
Requirement already satisfied: requests in /usr/local/lib/python3.11/dist-packages (from transformers==4.36.2) (2.32.3)
Collecting tokenizers<0.19,>=0.14 (from transformers==4.36.2)
  Downloading tokenizers-0.15.2-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (6.7 kB)
Requirement already satisfied: safetensors>=0.3.1 in /usr/local/lib/python3.11/dist-packages (from transformers==4.36.2) (0.5.3)
Requirement already satisfied: tqdm>=4.27 in /usr/local/lib/python3.11/dist-packages (from transformers==4.36.2) (4.67.1)
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)
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<=2025.3.0,>=2023.1.0 (from fsspec[http]<=2025.3.0,>=2023.1.0->datasets)
  Downloading fsspec-2025.3.0-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: scipy>=1.6.0 in /usr/local/lib/python3.11/dist-packages (from scikit-learn) (1.15.2)
Requirement already satisfied: joblib>=1.2.0 in /usr/local/lib/python3.11/dist-packages (from scikit-learn) (1.4.2)
Requirement already satisfied: threadpoolctl>=3.1.0 in /usr/local/lib/python3.11/dist-packages (from scikit-learn) (3.6.0)
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.11/dist-packages (from pandas) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.11/dist-packages (from pandas) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.11/dist-packages (from pandas) (2025.2)
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) (1.6.0)
Requirement already satisfied: multidict<7.0,>=4.5 in /usr/local/lib/python3.11/dist-packages (from aiohttp->datasets) (6.4.3)
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) (1.20.0)
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.11/dist-packages (from huggingface-hub<1.0,>=0.19
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.8.2->pandas) (1.17.0)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.11/dist-packages (from requests->transformers==4.36
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-packages (from requests->transformers==4.36.2) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-packages (from requests->transformers==4.36.2) (2
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.11/dist-packages (from requests->transformers==4.36.2) (2
Downloading transformers-4.36.2-py3-none-any.whl (8.2 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 8.2/8.2 MB 61.7 MB/s eta 0:00:00
Downloading datasets-3.5.1-py3-none-any.whl (491 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 491.4/491.4 kB 29.2 MB/s eta 0:00:00
Downloading dill-0.3.8-py3-none-any.whl (116 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 116.3/116.3 kB 10.7 MB/s eta 0:00:00
Downloading fsspec-2025.3.0-py3-none-any.whl (193 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 193.6/193.6 kB 16.1 MB/s eta 0:00:00
Downloading multiprocess-0.70.16-py311-none-any.whl (143 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 143.5/143.5 kB 13.4 MB/s eta 0:00:00
```

```
from datasets import load_dataset

dataset = load_dataset("go_emotions", "simplified")
```

```
/usr/local/lib/python3.11/dist-packages/huggingface_hub/utils/_auth.py:94: UserWarning:
The secret `HF_TOKEN` does not exist in your Colab secrets.
To authenticate with the Hugging Face Hub, create a token in your settings tab (https://huggingface.co/settings/tokens), set it as secre
You will be able to reuse this secret in all of your notebooks.
Please note that authentication is recommended but still optional to access public models or datasets.
  warnings.warn(
README.md: 100%                                                 9.40k/9.40k [00:00<00:00, 693kB/s]
train-00000-of-00001.parquet: 100%                              2.77M/2.77M [00:00<00:00, 12.4MB/s]
validation-00000-of-00001.parquet: 100%                         350k/350k [00:00<00:00, 11.7MB/s]
test-00000-of-00001.parquet: 100%                               347k/347k [00:00<00:00, 9.13MB/s]
Generating train split: 100%                                    43410/43410 [00:00<00:00, 338598.18 examples/s]
Generating validation split: 100%                               5426/5426 [00:00<00:00, 101974.65 examples/s]
Generating test split: 100%                                     5427/5427 [00:00<00:00, 159052.55 examples/s]
```

```python
def simplify_labels(example):
    example["labels"] = example["labels"][0] if isinstance(example["labels"], list) else example["labels"]
    return example

dataset = dataset.map(simplify_labels)
```

```
Map: 100%                                                       43410/43410 [00:06<00:00, 7534.12 examples/s]
Map: 100%                                                       5426/5426 [00:00<00:00, 8959.52 examples/s]
Map: 100%                                                       5427/5427 [00:00<00:00, 19591.19 examples/s]
```

```python
label_names = [
    'admiration', 'amusement', 'anger', 'annoyance', 'approval', 'caring',
    'confusion', 'curiosity', 'desire', 'disappointment', 'disapproval',
    'disgust', 'embarrassment', 'excitement', 'fear', 'gratitude', 'grief',
    'joy', 'love', 'nervousness', 'optimism', 'pride', 'realization', 'relief',
    'remorse', 'sadness', 'surprise', 'neutral'
]
id2label = {i: name for i, name in enumerate(label_names)}
label2id = {name: i for i, name in enumerate(label_names)}

example = dataset["train"][0]
label_id = example["labels"]
print(example)
print("Label name:", id2label[label_id])
```

```
{'text': "My favourite food is anything I didn't have to cook myself.", 'labels': 27, 'id': 'eebbqej'}
Label name: neutral
```

```python
from transformers import DistilBertTokenizerFast

tokenizer = DistilBertTokenizerFast.from_pretrained('distilbert-base-uncased')

def tokenize(batch):
    return tokenizer(batch["text"], truncation=True)

dataset = dataset.map(tokenize, batched=True)
```

```
/usr/local/lib/python3.11/dist-packages/transformers/utils/generic.py:441: FutureWarning: `torch.utils._pytree._register_pytree_node` is
  _torch_pytree._register_pytree_node(
The cache for model files in Transformers v4.22.0 has been updated. Migrating your old cache. This is a one-time only operation. You can
0/0 [00:00<?, ?it/s]
/usr/local/lib/python3.11/dist-packages/huggingface_hub/file_download.py:896: FutureWarning: `resume_download` is deprecated and will be
  warnings.warn(
tokenizer_config.json: 100%  48.0/48.0 [00:00<00:00, 3.09kB/s]
vocab.txt: 100%  232k/232k [00:00<00:00, 3.94MB/s]
tokenizer.json: 100%  466k/466k [00:00<00:00, 7.38MB/s]
config.json: 100%  483/483 [00:00<00:00, 51.6kB/s]
Map: 100%  43410/43410 [00:04<00:00, 8933.47 examples/s]
Map: 100%  5426/5426 [00:00<00:00, 14724.86 examples/s]
Map: 100%  5427/5427 [00:00<00:00, 15170.37 examples/s]
```

```python
from torch.utils.data import DataLoader
from transformers import DataCollatorWithPadding
import torch

for split in dataset:
    dataset[split].set_format(type="torch", columns=["input_ids", "attention_mask", "labels"])

data_collator = DataCollatorWithPadding(tokenizer=tokenizer, return_tensors="pt")

train_loader = DataLoader(dataset["train"], batch_size=8, shuffle=True, collate_fn=data_collator)
val_loader = DataLoader(dataset["validation"], batch_size=8, shuffle=False, collate_fn=data_collator)

```

```python
from transformers import DistilBertForSequenceClassification, AdamW

num_labels = len(label_names)
model = DistilBertForSequenceClassification.from_pretrained(
    "distilbert-base-uncased",
    num_labels=num_labels,
    id2label=id2label,
    label2id=label2id
)

device = torch.device("cuda" if torch.cuda.is_available() else "cpu")
model.to(device)
```

```
/usr/local/lib/python3.11/dist-packages/transformers/utils/generic.py:309: FutureWarning: `torch.utils._pytree._register_pytree_node` is
  _torch_pytree._register_pytree_node(
/usr/local/lib/python3.11/dist-packages/transformers/utils/generic.py:309: FutureWarning: `torch.utils._pytree._register_pytree_node` is
  _torch_pytree._register_pytree_node(
/usr/local/lib/python3.11/dist-packages/huggingface_hub/file_download.py:896: FutureWarning: `resume_download` is deprecated and will be
  warnings.warn(
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 [00:01<00:00, 203MB/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.
DistilBertForSequenceClassification(
  (distilbert): DistilBertModel(
    (embeddings): Embeddings(
      (word_embeddings): Embedding(30522, 768, padding_idx=0)
      (position_embeddings): Embedding(512, 768)
      (LayerNorm): LayerNorm((768,), eps=1e-12, elementwise_affine=True)
```

```python
optimizer = AdamW(model.parameters(), lr=5e-5)

model.train()
for epoch in range(8):
    total_loss = 0
    for batch in train_loader:
        batch = {k: v.to(device) for k, v in batch.items()}

        optimizer.zero_grad()
        outputs = model(**batch)
        loss = outputs.loss
        loss.backward()
        optimizer.step()

        total_loss += loss.item()

    avg_loss = total_loss / len(train_loader)
    print(f"Epoch {epoch+1} Loss: {avg_loss:.4f}")
```

```
/usr/local/lib/python3.11/dist-packages/transformers/optimization.py:429: FutureWarning: This implementation of AdamW is deprecated and
  warnings.warn(
You're using a DistilBertTokenizerFast tokenizer. Please note that with a fast tokenizer, using the `__call__` method is faster than usi
Epoch 1 Loss: 1.6416
Epoch 2 Loss: 1.2902
Epoch 3 Loss: 1.0155
Epoch 4 Loss: 0.7378
Epoch 5 Loss: 0.5197
Epoch 6 Loss: 0.3823
Epoch 7 Loss: 0.2966
Epoch 8 Loss: 0.2484
```

```python
model.eval()

def predict_emotion(text):
    inputs = tokenizer(text, return_tensors="pt", truncation=True, padding=True).to(device)
    with torch.no_grad():
        outputs = model(**inputs)
        logits = outputs.logits
        predicted_id = torch.argmax(logits, dim=1).item()
    return id2label[predicted_id]

print(predict_emotion("I failed my exam and feel terrible"))
print(predict_emotion("I miss my ex girlfriend"))
print(predict_emotion("I feel so disconnected and confused today"))
print(predict_emotion("I'm proud of what I accomplished this week"))
print(predict_emotion("Nothing really matters. What's the point?"))
```

```
sadness
sadness
confusion
admiration
```