Starter Notebook

Install and import required libraries

```
!pip install transformers datasets evaluate accelerate peft trl
bitsandbytes
!pip install nvidia-ml-py3
!pip install scikit-learn
```

```
Requirement already satisfied: transformers in
/usr/local/lib/python3.11/dist-packages (4.51.3)
Collecting datasets
  Downloading datasets-3.5.0-py3-none-any.whl.metadata (19 kB)
Collecting evaluate
  Downloading evaluate-0.4.3-py3-none-any.whl.metadata (9.2 kB)
Requirement already satisfied: accelerate in
/usr/local/lib/python3.11/dist-packages (1.5.2)
Requirement already satisfied: peft in /usr/local/lib/python3.11/dist-
packages (0.14.0)
Collecting trl
  Downloading trl-0.16.1-py3-none-any.whl.metadata (12 kB)
Collecting bitsandbytes
  Downloading bitsandbytes-0.45.5-py3-none-
manylinux_2_24_x86_64.whl.metadata (5.0 kB)
Requirement already satisfied: filelock in
/usr/local/lib/python3.11/dist-packages (from transformers) (3.18.0)
Requirement already satisfied: huggingface-hub<1.0,>=0.30.0 in
/usr/local/lib/python3.11/dist-packages (from transformers) (0.30.2)
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: 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)
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: psutil in
/usr/local/lib/python3.11/dist-packages (from accelerate) (5.9.5)
Requirement already satisfied: torch>=2.0.0 in
/usr/local/lib/python3.11/dist-packages (from accelerate)
(2.6.0+cu124)
Requirement already satisfied: rich in /usr/local/lib/python3.11/dist-
packages (from trl) (13.9.4)
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)
(6.4.3)
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)
(1.19.0)
Requirement already satisfied: typing-extensions>=3.7.4.3 in
/usr/local/lib/python3.11/dist-packages (from huggingface-
hub<1.0,>=0.30.0->transformers) (4.13.2)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.11/dist-packages (from requests->transformers)
(3.4.1)
```

```
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)
Requirement already satisfied: networkx in
/usr/local/lib/python3.11/dist-packages (from torch>=2.0.0-
>accelerate) (3.4.2)
Requirement already satisfied: jinja2 in
/usr/local/lib/python3.11/dist-packages (from torch>=2.0.0-
>accelerate) (3.1.6)
Collecting nvidia-cuda-nvrtc-cu12==12.4.127 (from torch>=2.0.0-
>accelerate)
  Downloading nvidia_cuda_nvrtc_cu12-12.4.127-py3-none-
manylinux2014_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-cuda-runtime-cu12==12.4.127 (from torch>=2.0.0-
>accelerate)
  Downloading nvidia_cuda_runtime_cu12-12.4.127-py3-none-
manylinux2014_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-cuda-cupti-cu12==12.4.127 (from torch>=2.0.0-
>accelerate)
  Downloading nvidia_cuda_cupti_cu12-12.4.127-py3-none-
manylinux2014_x86_64.whl.metadata (1.6 kB)
Collecting nvidia-cudnn-cu12==9.1.0.70 (from torch>=2.0.0->accelerate)
  Downloading nvidia_cudnn_cu12-9.1.0.70-py3-none-
manylinux2014_x86_64.whl.metadata (1.6 kB)
Collecting nvidia-cublas-cu12==12.4.5.8 (from torch>=2.0.0-
>accelerate)
  Downloading nvidia_cublas_cu12-12.4.5.8-py3-none-
manylinux2014_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-cufft-cu12==11.2.1.3 (from torch>=2.0.0->accelerate)
  Downloading nvidia_cufft_cu12-11.2.1.3-py3-none-
manylinux2014_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-curand-cu12==10.3.5.147 (from torch>=2.0.0-
>accelerate)
  Downloading nvidia_curand_cu12-10.3.5.147-py3-none-
manylinux2014_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-cusolver-cu12==11.6.1.9 (from torch>=2.0.0-
>accelerate)
  Downloading nvidia_cusolver_cu12-11.6.1.9-py3-none-
manylinux2014_x86_64.whl.metadata (1.6 kB)
Collecting nvidia-cusparse-cu12==12.3.1.170 (from torch>=2.0.0-
>accelerate)
  Downloading nvidia_cusparse_cu12-12.3.1.170-py3-none-
manylinux2014_x86_64.whl.metadata (1.6 kB)
```

```
Requirement already satisfied: nvidia-cusparselt-cu12==0.6.2 in
/usr/local/lib/python3.11/dist-packages (from torch>=2.0.0-
>accelerate) (0.6.2)
Requirement already satisfied: nvidia-nccl-cu12==2.21.5 in
/usr/local/lib/python3.11/dist-packages (from torch>=2.0.0-
>accelerate) (2.21.5)
Requirement already satisfied: nvidia-nvtx-cu12==12.4.127 in
/usr/local/lib/python3.11/dist-packages (from torch>=2.0.0-
>accelerate) (12.4.127)
Collecting nvidia-nvjitlink-cu12==12.4.127 (from torch>=2.0.0-
>accelerate)
  Downloading nvidia_nvjitlink_cu12-12.4.127-py3-none-
manylinux2014_x86_64.whl.metadata (1.5 kB)
Requirement already satisfied: triton==3.2.0 in
/usr/local/lib/python3.11/dist-packages (from torch>=2.0.0-
>accelerate) (3.2.0)
Requirement already satisfied: sympy==1.13.1 in
/usr/local/lib/python3.11/dist-packages (from torch>=2.0.0-
>accelerate) (1.13.1)
Requirement already satisfied: mpmath<1.4,>=1.1.0 in
/usr/local/lib/python3.11/dist-packages (from sympy==1.13.1-
>torch>=2.0.0->accelerate) (1.3.0)
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: markdown-it-py>=2.2.0 in
/usr/local/lib/python3.11/dist-packages (from rich->trl) (3.0.0)
Requirement already satisfied: pygments<3.0.0,>=2.13.0 in
/usr/local/lib/python3.11/dist-packages (from rich->trl) (2.18.0)
Requirement already satisfied: mdurl~=0.1 in
/usr/local/lib/python3.11/dist-packages (from markdown-it-py>=2.2.0-
>rich->trl) (0.1.2)
Requirement already satisfied: six>=1.5 in
/usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.8.2-
>pandas->datasets) (1.17.0)
Requirement already satisfied: MarkupSafe>=2.0 in
/usr/local/lib/python3.11/dist-packages (from jinja2->torch>=2.0.0-
>accelerate) (3.0.2)
Downloading datasets-3.5.0-py3-none-any.whl (491 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 491.2/491.2 kB 15.1 MB/s eta
0:00:00
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 84.0/84.0 kB 5.2 MB/s eta
0:00:00
```

```
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 336.4/336.4 kB 25.6 MB/s eta
0:00:00
anylinux_2_24_x86_64.whl (76.1 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 76.1/76.1 MB 10.8 MB/s eta
0:00:00
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 116.3/116.3 kB 6.2 MB/s eta
0:00:00
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 183.9/183.9 kB 13.1 MB/s eta
0:00:00
ultiprocess-0.70.16-py311-none-any.whl (143 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 143.5/143.5 kB 10.1 MB/s eta
0:00:00
anylinux2014_x86_64.whl (363.4 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 363.4/363.4 MB 2.7 MB/s eta
0:00:00
anylinux2014_x86_64.whl (13.8 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 13.8/13.8 MB 30.2 MB/s eta
0:00:00
anylinux2014_x86_64.whl (24.6 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 24.6/24.6 MB 40.1 MB/s eta
0:00:00
e_cu12-12.4.127-py3-none-manylinux2014_x86_64.whl (883 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 883.7/883.7 kB 31.5 MB/s eta
0:00:00
anylinux2014_x86_64.whl (664.8 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 664.8/664.8 MB 3.0 MB/s eta
0:00:00
anylinux2014_x86_64.whl (211.5 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 211.5/211.5 MB 5.4 MB/s eta
0:00:00
anylinux2014_x86_64.whl (56.3 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 56.3/56.3 MB 13.1 MB/s eta
0:00:00
anylinux2014_x86_64.whl (127.9 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 127.9/127.9 MB 7.5 MB/s eta
0:00:00
anylinux2014_x86_64.whl (207.5 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 207.5/207.5 MB 5.2 MB/s eta
0:00:00
anylinux2014_x86_64.whl (21.1 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 21.1/21.1 MB 81.0 MB/s eta
0:00:00
anylinux_2_17_x86_64.manylinux2014_x86_64.whl (194 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 194.8/194.8 kB 17.9 MB/s eta
0:00:00
e-cu12, nvidia-cuda-nvrtc-cu12, nvidia-cuda-cupti-cu12, nvidia-cublas-
cu12, fsspec, dill, nvidia-cusparse-cu12, nvidia-cudnn-cu12,
multiprocess, nvidia-cusolver-cu12, datasets, evaluate, bitsandbytes,
trl
```

```
  Attempting uninstall: nvidia-nvjitlink-cu12
    Found existing installation: nvidia-nvjitlink-cu12 12.5.82
    Uninstalling nvidia-nvjitlink-cu12-12.5.82:
      Successfully uninstalled nvidia-nvjitlink-cu12-12.5.82
  Attempting uninstall: nvidia-curand-cu12
    Found existing installation: nvidia-curand-cu12 10.3.6.82
    Uninstalling nvidia-curand-cu12-10.3.6.82:
      Successfully uninstalled nvidia-curand-cu12-10.3.6.82
  Attempting uninstall: nvidia-cufft-cu12
    Found existing installation: nvidia-cufft-cu12 11.2.3.61
    Uninstalling nvidia-cufft-cu12-11.2.3.61:
      Successfully uninstalled nvidia-cufft-cu12-11.2.3.61
  Attempting uninstall: nvidia-cuda-runtime-cu12
    Found existing installation: nvidia-cuda-runtime-cu12 12.5.82
    Uninstalling nvidia-cuda-runtime-cu12-12.5.82:
      Successfully uninstalled nvidia-cuda-runtime-cu12-12.5.82
  Attempting uninstall: nvidia-cuda-nvrtc-cu12
    Found existing installation: nvidia-cuda-nvrtc-cu12 12.5.82
    Uninstalling nvidia-cuda-nvrtc-cu12-12.5.82:
      Successfully uninstalled nvidia-cuda-nvrtc-cu12-12.5.82
  Attempting uninstall: nvidia-cuda-cupti-cu12
    Found existing installation: nvidia-cuda-cupti-cu12 12.5.82
    Uninstalling nvidia-cuda-cupti-cu12-12.5.82:
      Successfully uninstalled nvidia-cuda-cupti-cu12-12.5.82
  Attempting uninstall: nvidia-cublas-cu12
    Found existing installation: nvidia-cublas-cu12 12.5.3.2
    Uninstalling nvidia-cublas-cu12-12.5.3.2:
      Successfully uninstalled nvidia-cublas-cu12-12.5.3.2
  Attempting uninstall: fsspec
    Found existing installation: fsspec 2025.3.2
    Uninstalling fsspec-2025.3.2:
      Successfully uninstalled fsspec-2025.3.2
  Attempting uninstall: nvidia-cusparse-cu12
    Found existing installation: nvidia-cusparse-cu12 12.5.1.3
    Uninstalling nvidia-cusparse-cu12-12.5.1.3:
      Successfully uninstalled nvidia-cusparse-cu12-12.5.1.3
  Attempting uninstall: nvidia-cudnn-cu12
    Found existing installation: nvidia-cudnn-cu12 9.3.0.75
    Uninstalling nvidia-cudnn-cu12-9.3.0.75:
      Successfully uninstalled nvidia-cudnn-cu12-9.3.0.75
  Attempting uninstall: nvidia-cusolver-cu12
    Found existing installation: nvidia-cusolver-cu12 11.6.3.83
    Uninstalling nvidia-cusolver-cu12-11.6.3.83:
      Successfully uninstalled nvidia-cusolver-cu12-11.6.3.83
ERROR: pip's dependency resolver does not currently take into account
all the packages that are installed. This behaviour is the source of
the following dependency conflicts.
gcsfs 2025.3.2 requires fsspec==2025.3.2, but you have fsspec
2024.12.0 which is incompatible.
```

```
Successfully installed bitsandbytes-0.45.5 datasets-3.5.0 dill-0.3.8
evaluate-0.4.3 fsspec-2024.12.0 multiprocess-0.70.16 nvidia-cublas-
cu12-12.4.5.8 nvidia-cuda-cupti-cu12-12.4.127 nvidia-cuda-nvrtc-cu12-
12.4.127 nvidia-cuda-runtime-cu12-12.4.127 nvidia-cudnn-cu12-9.1.0.70
nvidia-cufft-cu12-11.2.1.3 nvidia-curand-cu12-10.3.5.147 nvidia-
cusolver-cu12-11.6.1.9 nvidia-cusparse-cu12-12.3.1.170 nvidia-
nvjitlink-cu12-12.4.127 trl-0.16.1 xxhash-3.5.0
Collecting nvidia-ml-py3
  Downloading nvidia-ml-py3-7.352.0.tar.gz (19 kB)
  Preparing metadata (setup.py) ... l-py3
  Building wheel for nvidia-ml-py3 (setup.py) ... l-py3:
filename=nvidia_ml_py3-7.352.0-py3-none-any.whl size=19172
sha256=caeb12e4756a8a16be4bed56d931ce8e48f81c4ebf0185c828dff2e6a1edae9
8
  Stored in directory:
/root/.cache/pip/wheels/47/50/9e/29dc79037d74c3c1bb4a8661fb608e8674b7e
4260d6a3f8f51
Successfully built nvidia-ml-py3
Installing collected packages: nvidia-ml-py3
Successfully installed nvidia-ml-py3-7.352.0
Requirement already satisfied: scikit-learn in
/usr/local/lib/python3.11/dist-packages (1.6.1)
Requirement already satisfied: numpy>=1.19.5 in
/usr/local/lib/python3.11/dist-packages (from scikit-learn) (2.0.2)
Requirement already satisfied: scipy>=1.6.0 in
/usr/local/lib/python3.11/dist-packages (from scikit-learn) (1.14.1)
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)

import os
import pandas as pd
import torch
from transformers import RobertaModel, RobertaTokenizer,
TrainingArguments, Trainer, DataCollatorWithPadding,
RobertaForSequenceClassification
from peft import LoraConfig, get_peft_model, PeftModel
from datasets import load_dataset, Dataset, ClassLabel
import pickle
```

Load Tokenizer and Preprocess Data

```
base_model = 'roberta-base'

dataset = load_dataset('ag_news', split='train')
tokenizer = RobertaTokenizer.from_pretrained(base_model)

def preprocess(examples):
```

```python
    tokenized = tokenizer(examples['text'], truncation=True,
padding=True)
    return tokenized

tokenized_dataset = dataset.map(preprocess, batched=True,
remove_columns=["text"])
tokenized_dataset = tokenized_dataset.rename_column("label", "labels")
```

```
/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
secret in your Google Colab and restart your session.
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(
```

{"model_id":"ac3ef2a4cbd54af49b4956e9419fd122","version_major":2,"version_minor":0}

{"model_id":"76360a417ad949ffbbc5483b1d6cba46","version_major":2,"version_minor":0}

{"model_id":"2191cef2956b468ebec4d04195710d95","version_major":2,"version_minor":0}

{"model_id":"5123ee2abf964f61ac25564ba2b28df3","version_major":2,"version_minor":0}

{"model_id":"9adac11d09e647fe95f734ed26694685","version_major":2,"version_minor":0}

{"model_id":"90d2dd39019f4f85ae62981cd91ffe8c","version_major":2,"version_minor":0}

{"model_id":"3894f9e0c0804a7fb77c1c62a2ba1b2e","version_major":2,"version_minor":0}

{"model_id":"69cc5130c71f4aa19f37a887497d49dd","version_major":2,"version_minor":0}

{"model_id":"a1dc69fc59634d19ace6ec2f3adcdaf3","version_major":2,"version_minor":0}

{"model_id":"9b6d85322b3e48e0ae9168676a559f3a","version_major":2,"version_minor":0}

{"model_id":"6640646067ac41d19dce32b4fba89aa3","version_major":2,"version_minor":0}

```
# Extract the number of classess and their names
num_labels = dataset.features['label'].num_classes
class_names = dataset.features["label"].names
print(f"number of labels: {num_labels}")
print(f"the labels: {class_names}")

# Create an id2label mapping
# We will need this for our classifier.
id2label = {i: label for i, label in enumerate(class_names)}

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

```
number of labels: 4
the labels: ['World', 'Sports', 'Business', 'Sci/Tech']
```

Load Pre-trained Model

Set up config for pretrained model and download it from hugging face

```
model = RobertaForSequenceClassification.from_pretrained(
    base_model,
    id2label=id2label)
model
```

```
Xet Storage is enabled for this repo, but the 'hf_xet' package is not
installed. Falling back to regular HTTP download. For better
performance, install the package with: `pip install
huggingface_hub[hf_xet]` or `pip install hf_xet`
WARNING:huggingface_hub.file_download:Xet Storage is enabled for this
repo, but the 'hf_xet' package is not installed. Falling back to
regular HTTP download. For better performance, install the package
with: `pip install huggingface_hub[hf_xet]` or `pip install hf_xet`
```

{"model_id":"74a68b127b80494ca460c219c4cd59b8","version_major":2,"vers
ion_minor":0}

```
Some weights of RobertaForSequenceClassification were not initialized
from the model checkpoint at roberta-base and are newly initialized:
['classifier.dense.bias', 'classifier.dense.weight',
'classifier.out_proj.bias', 'classifier.out_proj.weight']
You should probably TRAIN this model on a down-stream task to be able
to use it for predictions and inference.
```

```
RobertaForSequenceClassification(
  (roberta): RobertaModel(
    (embeddings): RobertaEmbeddings(
      (word_embeddings): Embedding(50265, 768, padding_idx=1)
      (position_embeddings): Embedding(514, 768, padding_idx=1)
      (token_type_embeddings): Embedding(1, 768)
```

```
      (LayerNorm): LayerNorm((768,), eps=1e-05,
elementwise_affine=True)
      (dropout): Dropout(p=0.1, inplace=False)
    )
    (encoder): RobertaEncoder(
      (layer): ModuleList(
        (0-11): 12 x RobertaLayer(
          (attention): RobertaAttention(
            (self): RobertaSdpaSelfAttention(
              (query): Linear(in_features=768, out_features=768,
bias=True)
              (key): Linear(in_features=768, out_features=768,
bias=True)
              (value): Linear(in_features=768, out_features=768,
bias=True)
              (dropout): Dropout(p=0.1, inplace=False)
            )
            (output): RobertaSelfOutput(
              (dense): Linear(in_features=768, out_features=768,
bias=True)
              (LayerNorm): LayerNorm((768,), eps=1e-05,
elementwise_affine=True)
              (dropout): Dropout(p=0.1, inplace=False)
            )
          )
          (intermediate): RobertaIntermediate(
            (dense): Linear(in_features=768, out_features=3072,
bias=True)
            (intermediate_act_fn): GELUActivation()
          )
          (output): RobertaOutput(
            (dense): Linear(in_features=3072, out_features=768,
bias=True)
            (LayerNorm): LayerNorm((768,), eps=1e-05,
elementwise_affine=True)
            (dropout): Dropout(p=0.1, inplace=False)
          )
        )
      )
    )
  )
  (classifier): RobertaClassificationHead(
    (dense): Linear(in_features=768, out_features=768, bias=True)
    (dropout): Dropout(p=0.1, inplace=False)
    (out_proj): Linear(in_features=768, out_features=4, bias=True)
  )
)
```

Anything from here on can be modified

```
# Split the original training set
split_datasets = tokenized_dataset.train_test_split(test_size=640,
seed=42)
train_dataset = split_datasets['train']
eval_dataset = split_datasets['test']
```

Setup LoRA Config

Setup PEFT config and get peft model for finetuning

```
# # PEFT Config
# peft_config = LoraConfig(
#     r=2,
#     lora_alpha=4,
#     lora_dropout=0.05,
#     bias = 'none',
#     target_modules = ['query'],
#     task_type="SEQ_CLS",
# )

# Configure LoRA
peft_config = LoraConfig(
    r=7,  # LoRA rank
    lora_alpha=16,  # Alpha parameter for scaling
    lora_dropout=0.05, # Dropout probability for LoRA layers
    target_modules=["query", "key", "value"], # Apply LoRA to these
layers
    bias="none",  # Don't train bias parameters
    task_type="SEQ_CLS", # Specify the task type
)

peft_model = get_peft_model(model, peft_config)
peft_model
```

```
PeftModelForSequenceClassification(
  (base_model): LoraModel(
    (model): RobertaForSequenceClassification(
      (roberta): RobertaModel(
        (embeddings): RobertaEmbeddings(
          (word_embeddings): Embedding(50265, 768, padding_idx=1)
          (position_embeddings): Embedding(514, 768, padding_idx=1)
          (token_type_embeddings): Embedding(1, 768)
          (LayerNorm): LayerNorm((768,), eps=1e-05,
elementwise_affine=True)
          (dropout): Dropout(p=0.1, inplace=False)
        )
        (encoder): RobertaEncoder(
```

```
          (layer): ModuleList(
            (0-11): 12 x RobertaLayer(
              (attention): RobertaAttention(
                (self): RobertaSdpaSelfAttention(
                  (query): lora.Linear(
                    (base_layer): Linear(in_features=768,
out_features=768, bias=True)
                    (lora_dropout): ModuleDict(
                      (default): Dropout(p=0.05, inplace=False)
                    )
                    (lora_A): ModuleDict(
                      (default): Linear(in_features=768,
out_features=7, bias=False)
                    )
                    (lora_B): ModuleDict(
                      (default): Linear(in_features=7,
out_features=768, bias=False)
                    )
                    (lora_embedding_A): ParameterDict()
                    (lora_embedding_B): ParameterDict()
                    (lora_magnitude_vector): ModuleDict()
                  )
                  (key): lora.Linear(
                    (base_layer): Linear(in_features=768,
out_features=768, bias=True)
                    (lora_dropout): ModuleDict(
                      (default): Dropout(p=0.05, inplace=False)
                    )
                    (lora_A): ModuleDict(
                      (default): Linear(in_features=768,
out_features=7, bias=False)
                    )
                    (lora_B): ModuleDict(
                      (default): Linear(in_features=7,
out_features=768, bias=False)
                    )
                    (lora_embedding_A): ParameterDict()
                    (lora_embedding_B): ParameterDict()
                    (lora_magnitude_vector): ModuleDict()
                  )
                  (value): lora.Linear(
                    (base_layer): Linear(in_features=768,
out_features=768, bias=True)
                    (lora_dropout): ModuleDict(
                      (default): Dropout(p=0.05, inplace=False)
                    )
                    (lora_A): ModuleDict(
                      (default): Linear(in_features=768,
out_features=7, bias=False)
```

```
                    )
                    (lora_B): ModuleDict(
                      (default): Linear(in_features=7,
out_features=768, bias=False)
                    )
                    (lora_embedding_A): ParameterDict()
                    (lora_embedding_B): ParameterDict()
                    (lora_magnitude_vector): ModuleDict()
                  )
                  (dropout): Dropout(p=0.1, inplace=False)
                )
                (output): RobertaSelfOutput(
                  (dense): Linear(in_features=768, out_features=768,
bias=True)
                  (LayerNorm): LayerNorm((768,), eps=1e-05,
elementwise_affine=True)
                  (dropout): Dropout(p=0.1, inplace=False)
                )
              )
              (intermediate): RobertaIntermediate(
                (dense): Linear(in_features=768, out_features=3072,
bias=True)
                (intermediate_act_fn): GELUActivation()
              )
              (output): RobertaOutput(
                (dense): Linear(in_features=3072, out_features=768,
bias=True)
                (LayerNorm): LayerNorm((768,), eps=1e-05,
elementwise_affine=True)
                (dropout): Dropout(p=0.1, inplace=False)
              )
            )
          )
        )
      )
      (classifier): ModulesToSaveWrapper(
        (original_module): RobertaClassificationHead(
          (dense): Linear(in_features=768, out_features=768,
bias=True)
          (dropout): Dropout(p=0.1, inplace=False)
          (out_proj): Linear(in_features=768, out_features=4,
bias=True)
        )
        (modules_to_save): ModuleDict(
          (default): RobertaClassificationHead(
            (dense): Linear(in_features=768, out_features=768,
bias=True)
            (dropout): Dropout(p=0.1, inplace=False)
            (out_proj): Linear(in_features=768, out_features=4,
```

```
bias=True)
          )
        )
      )
    )
  )
)

# print("Trainable parameters:")
# for name, param in peft_model.named_parameters():
#     if param.requires_grad:
#         print(name)

print('PEFT Model')
peft_model.print_trainable_parameters()
```

```
PEFT Model
trainable params: 980,740 || all params: 125,629,448 || trainable%: 
0.7807
```

Training Setup

```
# To track evaluation accuracy during training
from sklearn.metrics import accuracy_score, precision_score, 
recall_score, f1_score

def compute_metrics(pred):
    labels = pred.label_ids
    preds = pred.predictions.argmax(-1)
    # Calculate accuracy
    accuracy = accuracy_score(labels, preds)
    return {
        'accuracy': accuracy
    }

# Setup Training args
output_dir = "results"
training_args = TrainingArguments(
    output_dir=output_dir,
    report_to=None,
    eval_strategy='steps',
    logging_steps=100,
    learning_rate=1e-5,
    num_train_epochs=1,
    max_steps=1200,
    use_cpu=False,
    dataloader_num_workers=4,
    per_device_train_batch_size=16,
    per_device_eval_batch_size=64,
    optim="adamw_torch",
```

```
    gradient_checkpointing=False,
    gradient_checkpointing_kwargs={'use_reentrant':True}
)

def get_trainer(model):
    return Trainer(
        model=model,
        args=training_args,
        compute_metrics=compute_metrics,
        train_dataset=train_dataset,
        eval_dataset=eval_dataset,
        data_collator=data_collator,
    )
```

Start Training

```
peft_lora_finetuning_trainer = get_trainer(peft_model)

result = peft_lora_finetuning_trainer.train()
```

```
No label_names provided for model class
`PeftModelForSequenceClassification`. Since `PeftModel` hides base
models input arguments, if label_names is not given, label_names can't
be set automatically within `Trainer`. Note that empty label_names
list will be used instead.
/usr/local/lib/python3.11/dist-packages/torch/utils/data/dataloader.py
:624: UserWarning: This DataLoader will create 4 worker processes in
total. Our suggested max number of worker in current system is 2,
which is smaller than what this DataLoader is going to create. Please
be aware that excessive worker creation might get DataLoader running
slow or even freeze, lower the worker number to avoid potential
slowness/freeze if necessary.
  warnings.warn(
wandb: WARNING The `run_name` is currently set to the same value as
`TrainingArguments.output_dir`. If this was not intended, please
specify a different run name by setting the
`TrainingArguments.run_name` parameter.
wandb: Using wandb-core as the SDK backend.  Please refer to
https://wandb.me/wandb-core for more information.

<IPython.core.display.Javascript object>

wandb: Logging into wandb.ai. (Learn how to deploy a W&B server
locally: https://wandb.me/wandb-server)
wandb: You can find your API key in your browser here:
https://wandb.ai/authorize
wandb: Paste an API key from your profile and hit enter:

 ..........
```

```
wandb: WARNING If you're specifying your api key in code, ensure this
code is not shared publicly.
wandb: WARNING Consider setting the WANDB_API_KEY environment
variable, or running `wandb login` from the command line.
wandb: No netrc file found, creating one.
wandb: Appending key for api.wandb.ai to your netrc file: /root/.netrc
wandb: Currently logged in as: dataencapsulate (dataencapsulate-new-
york-university) to https://api.wandb.ai. Use `wandb login --relogin`
to force relogin

<IPython.core.display.HTML object>

<IPython.core.display.HTML object>

<IPython.core.display.HTML object>

<IPython.core.display.HTML object>

<IPython.core.display.HTML object>

/usr/local/lib/python3.11/dist-packages/torch/utils/data/
dataloader.py:624: UserWarning: This DataLoader will create 4 worker
processes in total. Our suggested max number of worker in current
system is 2, which is smaller than what this DataLoader is going to
create. Please be aware that excessive worker creation might get
DataLoader running slow or even freeze, lower the worker number to
avoid potential slowness/freeze if necessary.
  warnings.warn(

<IPython.core.display.HTML object>

/usr/local/lib/python3.11/dist-packages/torch/utils/data/
dataloader.py:624: UserWarning: This DataLoader will create 4 worker
processes in total. Our suggested max number of worker in current
system is 2, which is smaller than what this DataLoader is going to
create. Please be aware that excessive worker creation might get
DataLoader running slow or even freeze, lower the worker number to
avoid potential slowness/freeze if necessary.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/torch/utils/data/dataloader.py
:624: UserWarning: This DataLoader will create 4 worker processes in
total. Our suggested max number of worker in current system is 2,
which is smaller than what this DataLoader is going to create. Please
be aware that excessive worker creation might get DataLoader running
slow or even freeze, lower the worker number to avoid potential
slowness/freeze if necessary.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/torch/utils/data/dataloader.py
:624: UserWarning: This DataLoader will create 4 worker processes in
total. Our suggested max number of worker in current system is 2,
which is smaller than what this DataLoader is going to create. Please
```

```
be aware that excessive worker creation might get DataLoader running
slow or even freeze, lower the worker number to avoid potential
slowness/freeze if necessary.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/torch/utils/data/dataloader.py
:624: UserWarning: This DataLoader will create 4 worker processes in
total. Our suggested max number of worker in current system is 2,
which is smaller than what this DataLoader is going to create. Please
be aware that excessive worker creation might get DataLoader running
slow or even freeze, lower the worker number to avoid potential
slowness/freeze if necessary.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/torch/utils/data/dataloader.py
:624: UserWarning: This DataLoader will create 4 worker processes in
total. Our suggested max number of worker in current system is 2,
which is smaller than what this DataLoader is going to create. Please
be aware that excessive worker creation might get DataLoader running
slow or even freeze, lower the worker number to avoid potential
slowness/freeze if necessary.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/torch/utils/data/dataloader.py
:624: UserWarning: This DataLoader will create 4 worker processes in
total. Our suggested max number of worker in current system is 2,
which is smaller than what this DataLoader is going to create. Please
be aware that excessive worker creation might get DataLoader running
slow or even freeze, lower the worker number to avoid potential
slowness/freeze if necessary.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/torch/utils/data/dataloader.py
:624: UserWarning: This DataLoader will create 4 worker processes in
total. Our suggested max number of worker in current system is 2,
which is smaller than what this DataLoader is going to create. Please
be aware that excessive worker creation might get DataLoader running
slow or even freeze, lower the worker number to avoid potential
slowness/freeze if necessary.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/torch/utils/data/dataloader.py
:624: UserWarning: This DataLoader will create 4 worker processes in
total. Our suggested max number of worker in current system is 2,
which is smaller than what this DataLoader is going to create. Please
be aware that excessive worker creation might get DataLoader running
slow or even freeze, lower the worker number to avoid potential
slowness/freeze if necessary.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/torch/utils/data/dataloader.py
:624: UserWarning: This DataLoader will create 4 worker processes in
total. Our suggested max number of worker in current system is 2,
which is smaller than what this DataLoader is going to create. Please
be aware that excessive worker creation might get DataLoader running
```

```
slow or even freeze, lower the worker number to avoid potential
slowness/freeze if necessary.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/torch/utils/data/dataloader.py
:624: UserWarning: This DataLoader will create 4 worker processes in
total. Our suggested max number of worker in current system is 2,
which is smaller than what this DataLoader is going to create. Please
be aware that excessive worker creation might get DataLoader running
slow or even freeze, lower the worker number to avoid potential
slowness/freeze if necessary.
  warnings.warn(
/usr/local/lib/python3.11/dist-packages/torch/utils/data/dataloader.py
:624: UserWarning: This DataLoader will create 4 worker processes in
total. Our suggested max number of worker in current system is 2,
which is smaller than what this DataLoader is going to create. Please
be aware that excessive worker creation might get DataLoader running
slow or even freeze, lower the worker number to avoid potential
slowness/freeze if necessary.
  warnings.warn(
```

Evaluate Finetuned Model

Performing Inference on Custom Input

Uncomment following functions for running inference on custom inputs

```
def classify(model, tokenizer, text):
    device = torch.device("cuda" if torch.cuda.is_available() else
"cpu")
    inputs = tokenizer(text, truncation=True, padding=True,
return_tensors="pt").to(device)
    output = model(**inputs)

    prediction = output.logits.argmax(dim=-1).item()

    print(f'\n Class: {prediction}, Label: {id2label[prediction]},
Text: {text}')
    return id2label[prediction]

classify( peft_model, tokenizer, "Kederis proclaims innocence Olympic
champion Kostas Kederis today left hospital ahead of his date with IOC
inquisitors claiming his ...")
classify( peft_model, tokenizer, "Wall St. Bears Claw Back Into the
Black (Reuters) Reuters - Short-sellers, Wall Street's dwindling\band
of ultra-cynics, are seeing green again.")
```

```
 Class: 1, Label: Sports, Text: Kederis proclaims innocence Olympic
champion Kostas Kederis today left hospital ahead of his date with IOC
```

```
inquisitors claiming his ...

 Class: 2, Label: Business, Text: Wall St. Bears Claw Back Into the
Black (Reuters) Reuters - Short-sellers, Wall Street's dwindlingand of
ultra-cynics, are seeing green again.
```

{"type":"string"}

Run Inference on eval_dataset

```python
from torch.utils.data import DataLoader
import evaluate
from tqdm import tqdm

def evaluate_model(inference_model, dataset, labelled=True,
batch_size=8, data_collator=None):
    """
    Evaluate a PEFT model on a dataset.

    Args:
        inference_model: The model to evaluate.
        dataset: The dataset (Hugging Face Dataset) to run inference
on.
        labelled (bool): If True, the dataset includes labels and
metrics will be computed.
                         If False, only predictions will be returned.
        batch_size (int): Batch size for inference.
        data_collator: Function to collate batches. If None, the
default collate_fn is used.

    Returns:
        If labelled is True, returns a tuple (metrics, predictions)
        If labelled is False, returns the predictions.
    """
    # Create the DataLoader
    eval_dataloader = DataLoader(dataset, batch_size=batch_size,
collate_fn=data_collator)
    device = torch.device("cuda" if torch.cuda.is_available() else
"cpu")

    inference_model.to(device)
    inference_model.eval()

    all_predictions = []
    if labelled:
        metric = evaluate.load('accuracy')

    # Loop over the DataLoader
    for batch in tqdm(eval_dataloader):
        # Move each tensor in the batch to the device
```

```
        batch = {k: v.to(device) for k, v in batch.items()}
        with torch.no_grad():
            outputs = inference_model(**batch)
        predictions = outputs.logits.argmax(dim=-1)
        all_predictions.append(predictions.cpu())

        if labelled:
            # Expecting that labels are provided under the "labels"
key.
            references = batch["labels"]
            metric.add_batch(
                predictions=predictions.cpu().numpy(),
                references=references.cpu().numpy()
            )

    # Concatenate predictions from all batches
    all_predictions = torch.cat(all_predictions, dim=0)

    if labelled:
        eval_metric = metric.compute()
        print("Evaluation Metric:", eval_metric)
        return eval_metric, all_predictions
    else:
        return all_predictions

# Check evaluation accuracy
_, _ = evaluate_model(peft_model, eval_dataset, True, 8,
data_collator)
```

{"model_id":"76ffaf23694c4125bd2130e44e3270f1","version_major":2,"vers
ion_minor":0}

```
100%|██████████| 80/80 [00:11<00:00,  6.91it/s]

Evaluation Metric: {'accuracy': 0.8890625}
```

Run Inference on unlabelled dataset

```
#Load your unlabelled data
unlabelled_dataset = pd.read_pickle("test_unlabelled.pkl")
test_dataset = unlabelled_dataset.map(preprocess, batched=True,
remove_columns=["text"])
unlabelled_dataset
```

{"model_id":"90af531e8f364d358a6db8a15088bb09","version_major":2,"vers
ion_minor":0}

```
Dataset({
    features: ['text'],
```

```
    num_rows: 8000
})
```

```python
# Run inference and save predictions
preds = evaluate_model(peft_model, test_dataset, False, 8,
data_collator)
df_output = pd.DataFrame({
    'ID': range(len(preds)),
    'Label': preds.numpy()  # or preds.tolist()
})
df_output.to_csv(os.path.join(output_dir,"inference_output.csv"),
index=False)
print("Inference complete. Predictions saved to inference_output.csv")
```

```
100%|██████████| 1000/1000 [01:51<00:00,  8.96it/s]
```

```
Inference complete. Predictions saved to inference_output.csv
```

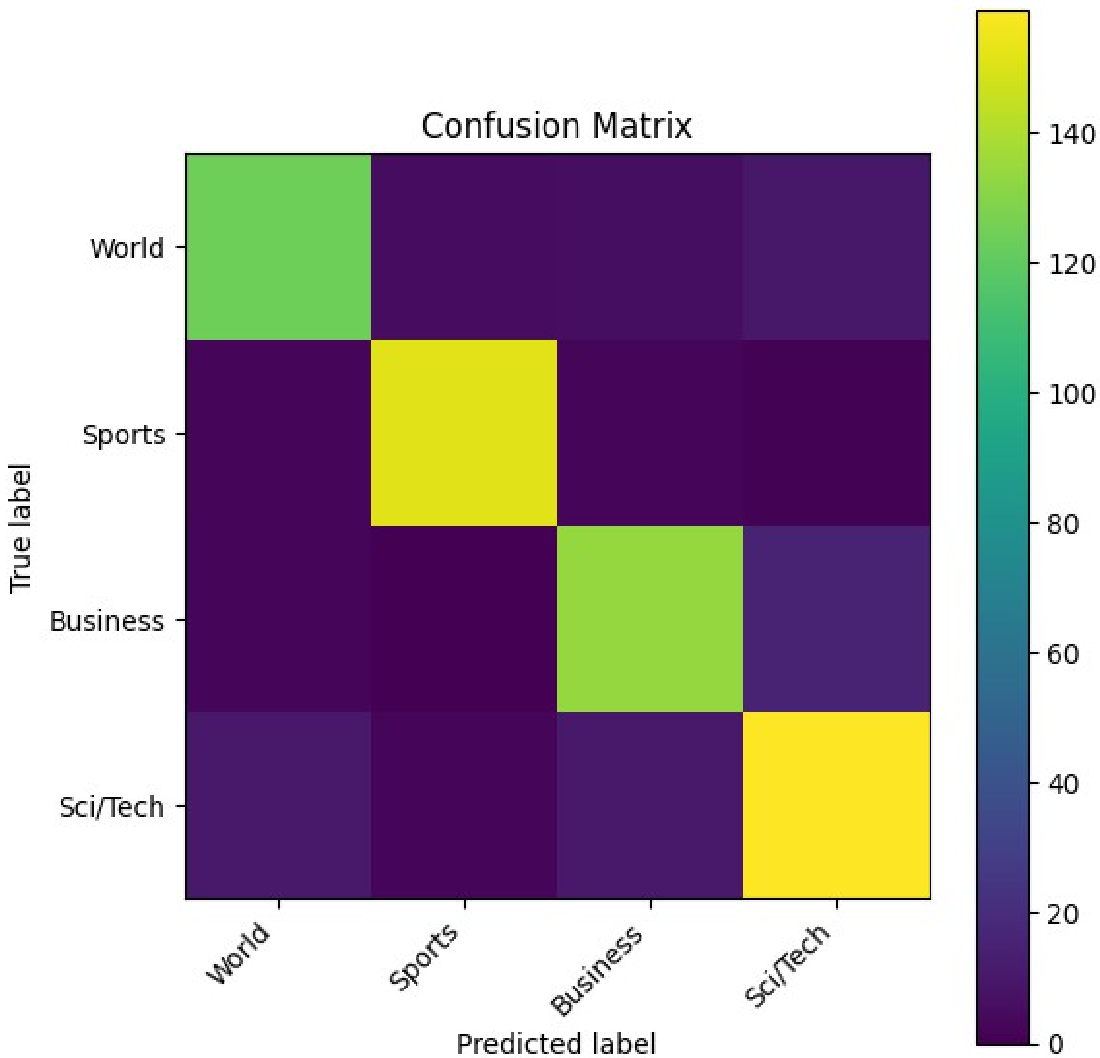
```python
# %% Cell 1: Compute predictions, overall & per-class metrics, and
class distribution
from sklearn.metrics import classification_report, confusion_matrix
import numpy as np
import pandas as pd

# 1. Run predictions on validation set
pred_output = peft_lora_finetuning_trainer.predict(eval_dataset)
preds = np.argmax(pred_output.predictions, axis=1)
labels = pred_output.label_ids

# 2. Overall performance + per-category metrics
report = classification_report(labels, preds,
                               target_names=list(id2label.values()),
                               output_dict=True)
report_df = pd.DataFrame(report).transpose()
print("\n=== Classification Report ===")
print(report_df)

# 3. Predicted class distribution
class_dist =
pd.Series(preds).value_counts().sort_index().rename_axis('class_id').r
eset_index(name='count')
class_dist['class_name'] = class_dist['class_id'].map(id2label)
print("\n=== Predicted Class Distribution ===")
print(class_dist)
```

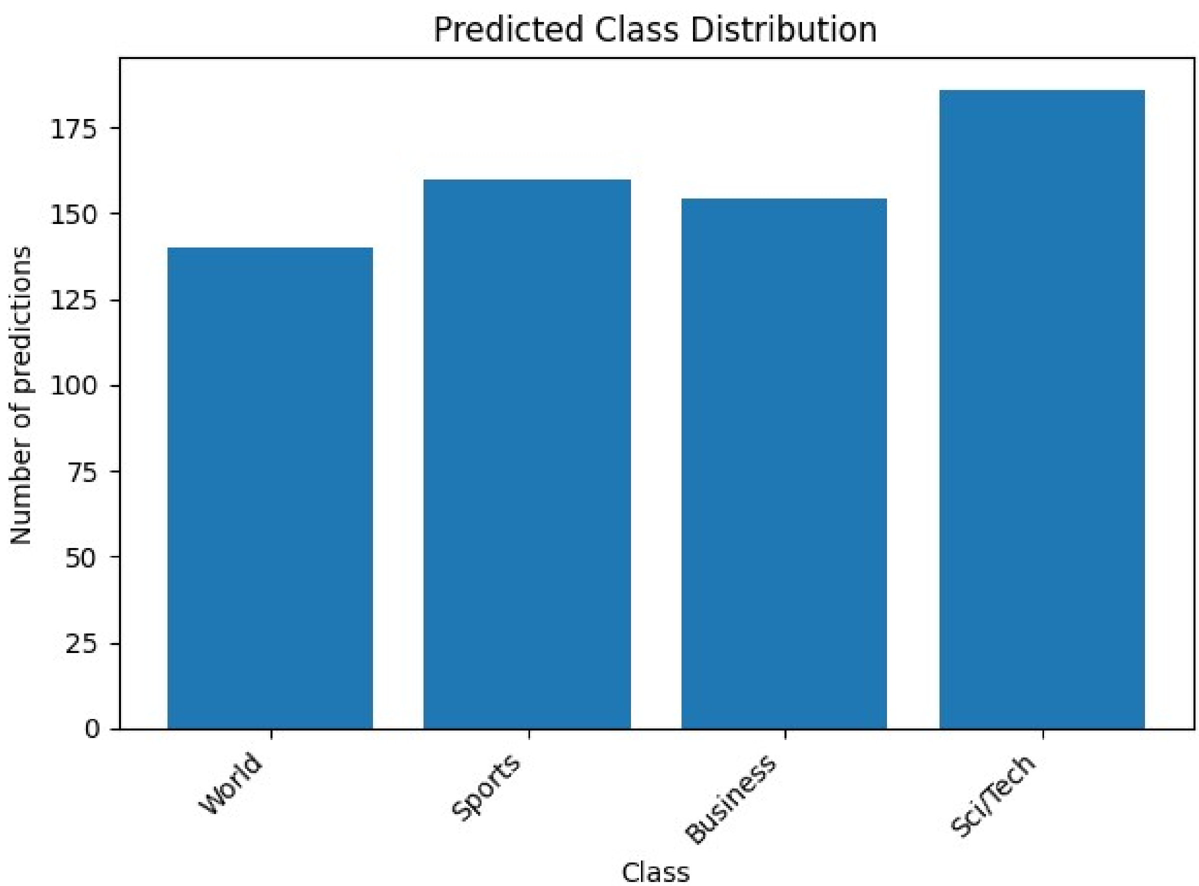
```
/usr/local/lib/python3.11/dist-packages/torch/utils/data/
dataloader.py:624: UserWarning: This DataLoader will create 4 worker
processes in total. Our suggested max number of worker in current
system is 2, which is smaller than what this DataLoader is going to
```

```
create. Please be aware that excessive worker creation might get
DataLoader running slow or even freeze, lower the worker number to
avoid potential slowness/freeze if necessary.
  warnings.warn(

<IPython.core.display.HTML object>


=== Classification Report ===
              precision    recall  f1-score     support
World          0.885714  0.855172  0.870175  145.000000
Sports         0.950000  0.962025  0.955975  158.000000
Business       0.870130  0.875817  0.872964  153.000000
Sci/Tech       0.854839  0.864130  0.859459  184.000000
accuracy       0.889062  0.889062  0.889062    0.889062
macro avg      0.890171  0.889286  0.889643  640.000000
weighted avg   0.888982  0.889062  0.888943  640.000000

=== Predicted Class Distribution ===
   class_id  count class_name
0         0    140      World
1         1    160     Sports
2         2    154   Business
3         3    186   Sci/Tech
```

```
# %% Cell 2: Confusion matrix & class-distribution bar chart
import matplotlib.pyplot as plt
import numpy as np
from sklearn.metrics import confusion_matrix

# Confusion matrix
conf_mat = confusion_matrix(labels, preds)
fig, ax = plt.subplots(figsize=(6,6))
im = ax.imshow(conf_mat, interpolation='nearest')
ax.set_xticks(np.arange(len(id2label)))
ax.set_yticks(np.arange(len(id2label)))
ax.set_xticklabels(list(id2label.values()), rotation=45, ha='right')
ax.set_yticklabels(list(id2label.values()))
ax.set_xlabel("Predicted label")
ax.set_ylabel("True label")
ax.set_title("Confusion Matrix")
fig.colorbar(im, ax=ax)
plt.tight_layout()
plt.show()

# Class distribution bar plot
fig, ax = plt.subplots()
ax.bar(class_dist['class_name'], class_dist['count'])
ax.set_xlabel("Class")
ax.set_ylabel("Number of predictions")
```

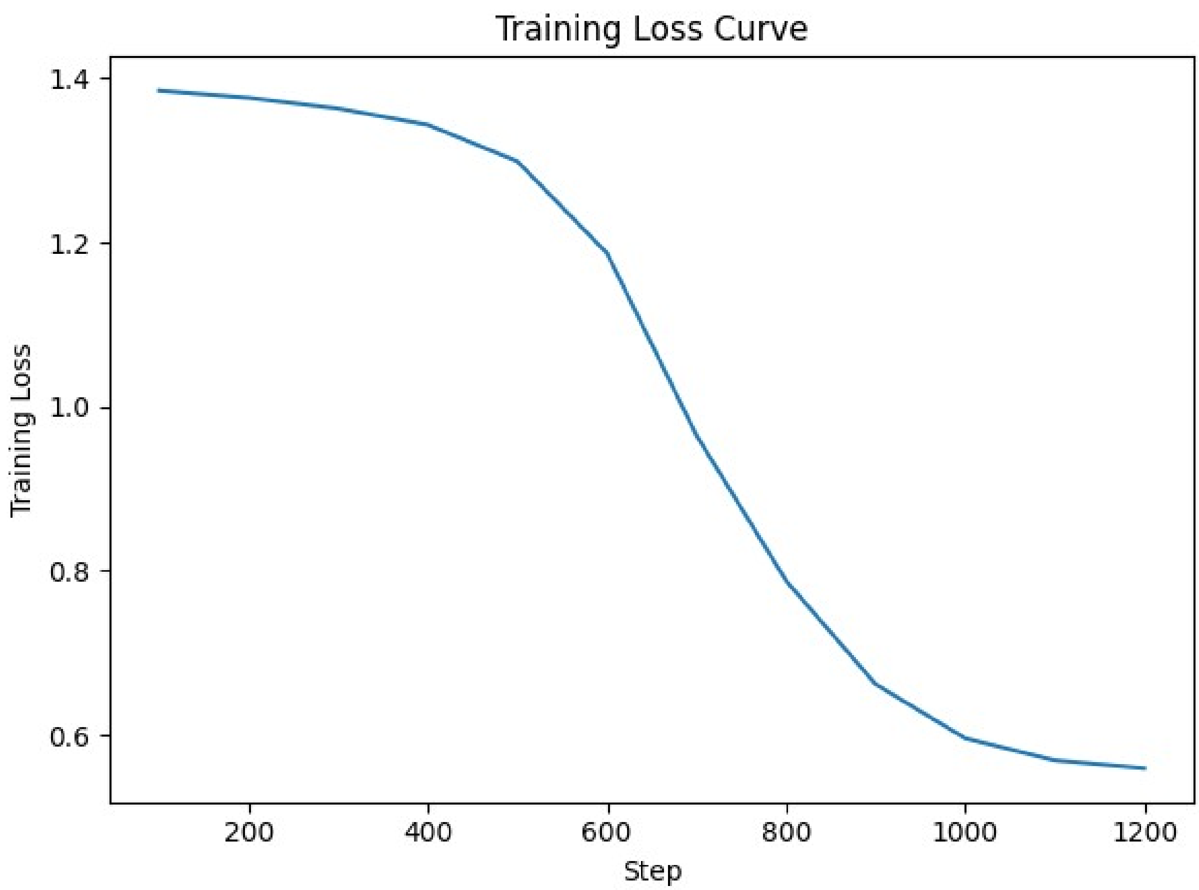
```
ax.set_title("Predicted Class Distribution")
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
plt.show()
```

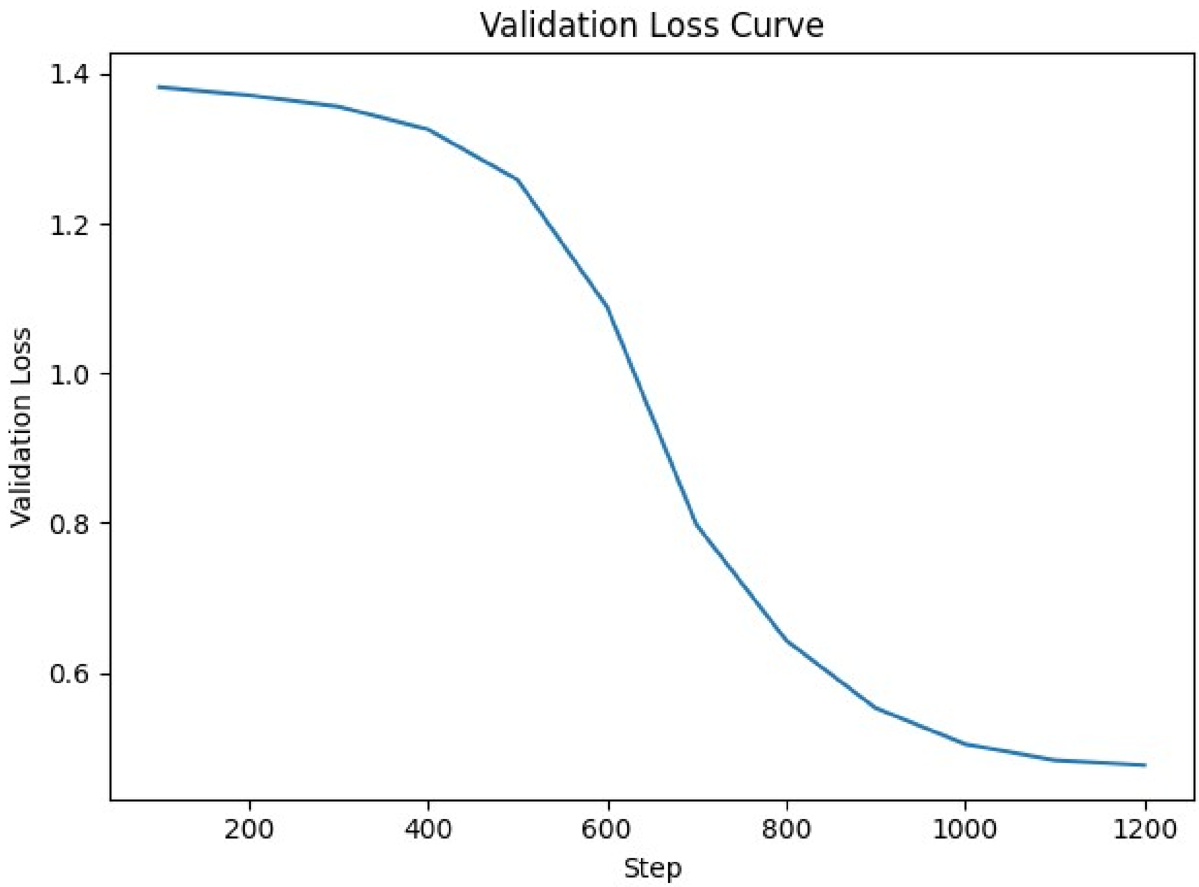


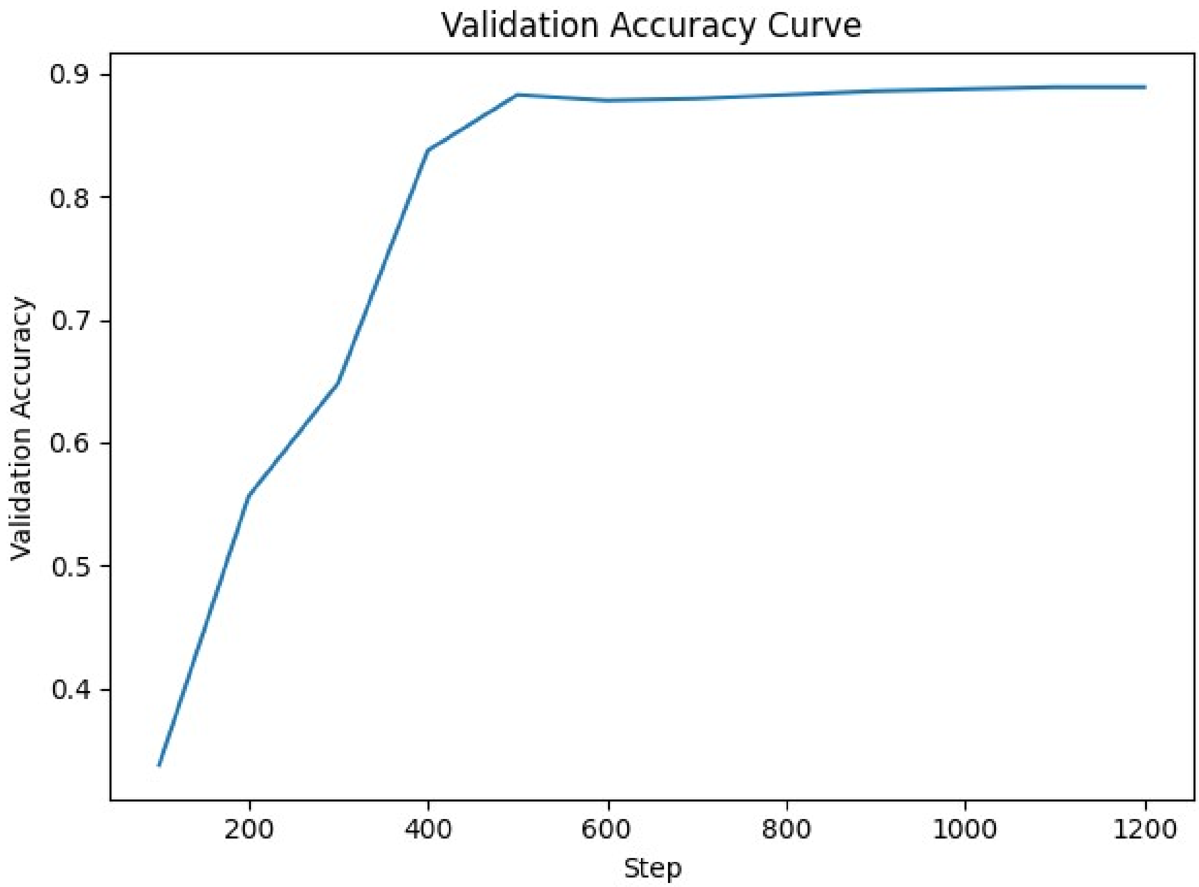



```
# %% Cell 3: Training & validation loss/accuracy curves
import matplotlib.pyplot as plt

# Extract from trainer state
history = peft_lora_finetuning_trainer.state.log_history

# Training loss entries (ones with 'loss' but no 'eval_loss')
train_steps  = [e['step'] for e in history if 'loss' in e and
'eval_loss' not in e]
train_losses = [e['loss'] for e in history if 'loss' in e and
'eval_loss' not in e]

# Validation metrics (those with 'eval_loss' and 'eval_accuracy')
eval_steps  = [e['step'] for e in history if 'eval_loss' in e]
eval_losses = [e['eval_loss'] for e in history if 'eval_loss' in e]
eval_accs   = [e['eval_accuracy'] for e in history if 'eval_accuracy'
in e]

# Plot training loss
plt.figure()
plt.plot(train_steps, train_losses)
plt.xlabel("Step")
```

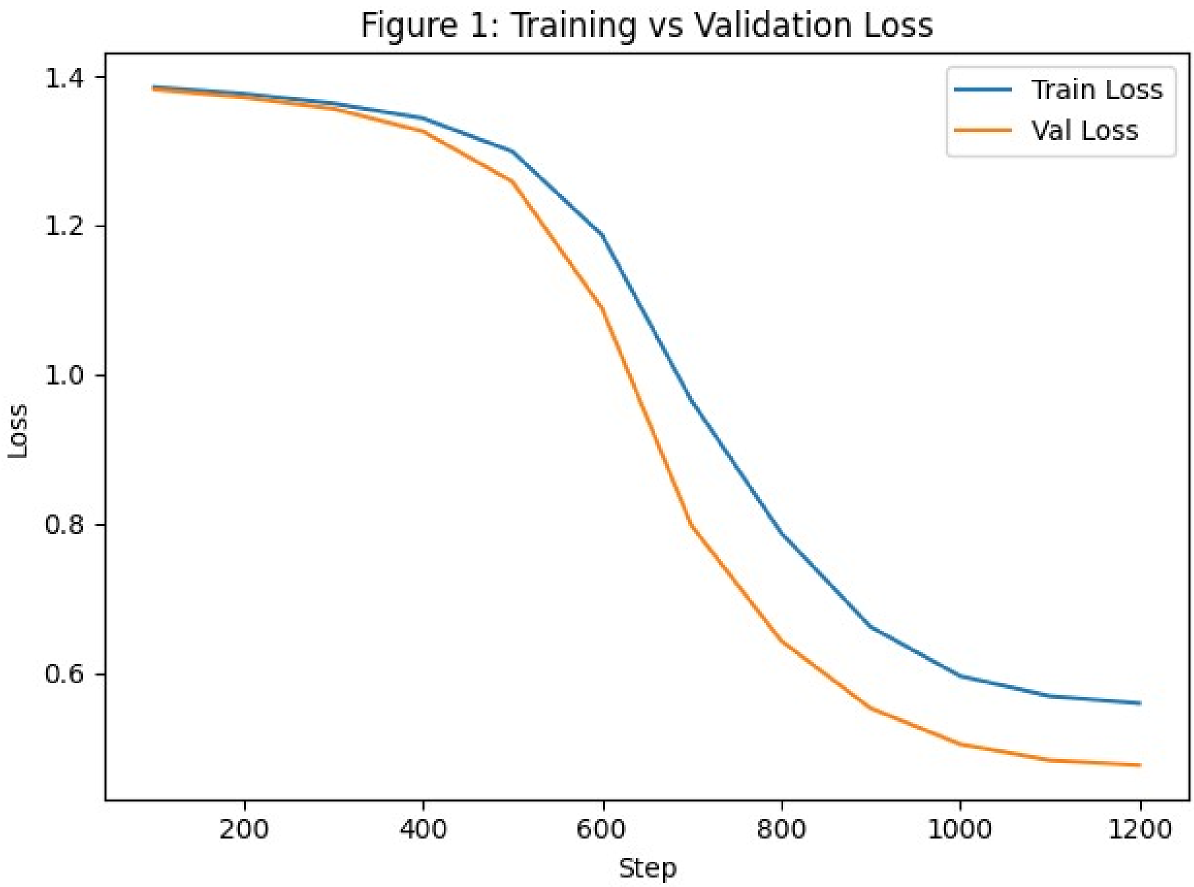
```
plt.ylabel("Training Loss")
plt.title("Training Loss Curve")
plt.tight_layout()
plt.show()

# Plot validation loss
plt.figure()
plt.plot(eval_steps, eval_losses)
plt.xlabel("Step")
plt.ylabel("Validation Loss")
plt.title("Validation Loss Curve")
plt.tight_layout()
plt.show()

# Plot validation accuracy
plt.figure()
plt.plot(eval_steps, eval_accs)
plt.xlabel("Step")
plt.ylabel("Validation Accuracy")
plt.title("Validation Accuracy Curve")
plt.tight_layout()
plt.show()
```




Validation Loss Curve
Validation Loss
1.4
1.2
1.0
0.8
0.6
200
400
600
800
1000
1200
Step




```
# %% Cell: Combined Loss Curve (Figure 1)
import matplotlib.pyplot as plt

# extract from your trainer state, as before
history = peft_lora_finetuning_trainer.state.log_history
train_steps  = [e['step']      for e in history if 'loss' in e and 'eval_loss' not in e]
train_losses = [e['loss']      for e in history if 'loss' in e and 'eval_loss' not in e]
eval_steps   = [e['step']      for e in history if 'eval_loss' in e]
eval_losses  = [e['eval_loss'] for e in history if 'eval_loss' in e]

plt.figure()
plt.plot(train_steps, train_losses, label='Train Loss')
plt.plot(eval_steps, eval_losses, label='Val Loss')
plt.xlabel('Step')
plt.ylabel('Loss')
plt.title('Figure 1: Training vs Validation Loss')
plt.legend()
plt.tight_layout()
plt.show()
```




```
# %% Cell: Figure 2 – Training vs Validation Accuracy
# 1) Get train & val accuracy in one pass each
train_pred = peft_lora_finetuning_trainer.predict(train_dataset)
val_pred   = peft_lora_finetuning_trainer.predict(eval_dataset)

# Depending on your HF version the metric key may be 'test_accuracy'
or 'eval_accuracy'
train_acc = train_pred.metrics.get('test_accuracy',
train_pred.metrics.get('eval_accuracy'))
val_acc   = val_pred.metrics.get('test_accuracy',
val_pred.metrics.get('eval_accuracy'))

# 2) Plot bar chart
import matplotlib.pyplot as plt

plt.figure()
plt.bar(['Train','Validation'], [train_acc, val_acc])
plt.ylim(0, 1.0)
plt.ylabel('Accuracy')
plt.title('Figure 2: Training vs Validation Accuracy')
plt.tight_layout()
plt.show()
```

```
/usr/local/lib/python3.11/dist-packages/torch/utils/data/
dataloader.py:624: UserWarning: This DataLoader will create 4 worker
processes in total. Our suggested max number of worker in current
system is 2, which is smaller than what this DataLoader is going to
create. Please be aware that excessive worker creation might get
DataLoader running slow or even freeze, lower the worker number to
avoid potential slowness/freeze if necessary.
  warnings.warn(

<IPython.core.display.HTML object>

/usr/local/lib/python3.11/dist-packages/torch/utils/data/
dataloader.py:624: UserWarning: This DataLoader will create 4 worker
processes in total. Our suggested max number of worker in current
system is 2, which is smaller than what this DataLoader is going to
create. Please be aware that excessive worker creation might get
DataLoader running slow or even freeze, lower the worker number to
avoid potential slowness/freeze if necessary.
  warnings.warn(

<IPython.core.display.HTML object>
```