```
!pip install datasets
```

```
Collecting datasets
  Downloading datasets-3.2.0-py3-none-any.whl.metadata (20 kB)
Requirement already satisfied: filelock in
/usr/local/lib/python3.10/dist-packages (from datasets) (3.16.1)
Requirement already satisfied: numpy>=1.17 in
/usr/local/lib/python3.10/dist-packages (from datasets) (1.26.4)
Requirement already satisfied: pyarrow>=15.0.0 in
/usr/local/lib/python3.10/dist-packages (from datasets) (17.0.0)
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/python3.10/dist-packages (from datasets) (2.2.2)
Requirement already satisfied: requests>=2.32.2 in
/usr/local/lib/python3.10/dist-packages (from datasets) (2.32.3)
Requirement already satisfied: tqdm>=4.66.3 in
/usr/local/lib/python3.10/dist-packages (from datasets) (4.67.1)
Collecting xxhash (from datasets)
  Downloading xxhash-3.5.0-cp310-cp310-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (12 kB)
Collecting multiprocess<0.70.17 (from datasets)
  Downloading multiprocess-0.70.16-py310-none-any.whl.metadata (7.2
kB)
Collecting fsspec<=2024.9.0,>=2023.1.0 (from
fsspec[http]<=2024.9.0,>=2023.1.0->datasets)
  Downloading fsspec-2024.9.0-py3-none-any.whl.metadata (11 kB)
Requirement already satisfied: aiohttp in
/usr/local/lib/python3.10/dist-packages (from datasets) (3.11.10)
Requirement already satisfied: huggingface-hub>=0.23.0 in
/usr/local/lib/python3.10/dist-packages (from datasets) (0.27.0)
Requirement already satisfied: packaging in
/usr/local/lib/python3.10/dist-packages (from datasets) (24.2)
Requirement already satisfied: pyyaml>=5.1 in
/usr/local/lib/python3.10/dist-packages (from datasets) (6.0.2)
Requirement already satisfied: aiohappyeyeballs>=2.3.0 in
/usr/local/lib/python3.10/dist-packages (from aiohttp->datasets)
(2.4.4)
Requirement already satisfied: aiosignal>=1.1.2 in
/usr/local/lib/python3.10/dist-packages (from aiohttp->datasets)
(1.3.2)
Requirement already satisfied: async-timeout<6.0,>=4.0 in
/usr/local/lib/python3.10/dist-packages (from aiohttp->datasets)
(4.0.3)
Requirement already satisfied: attrs>=17.3.0 in
/usr/local/lib/python3.10/dist-packages (from aiohttp->datasets)
(24.3.0)
Requirement already satisfied: frozenlist>=1.1.1 in
/usr/local/lib/python3.10/dist-packages (from aiohttp->datasets)
(1.5.0)
```

```
Requirement already satisfied: multidict<7.0,>=4.5 in
/usr/local/lib/python3.10/dist-packages (from aiohttp->datasets)
(6.1.0)
Requirement already satisfied: propcache>=0.2.0 in
/usr/local/lib/python3.10/dist-packages (from aiohttp->datasets)
(0.2.1)
Requirement already satisfied: yarl<2.0,>=1.17.0 in
/usr/local/lib/python3.10/dist-packages (from aiohttp->datasets)
(1.18.3)
Requirement already satisfied: typing-extensions>=3.7.4.3 in
/usr/local/lib/python3.10/dist-packages (from huggingface-hub>=0.23.0-
>datasets) (4.12.2)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.10/dist-packages (from requests>=2.32.2-
>datasets) (3.4.0)
Requirement already satisfied: idna<4,>=2.5 in
/usr/local/lib/python3.10/dist-packages (from requests>=2.32.2-
>datasets) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.10/dist-packages (from requests>=2.32.2-
>datasets) (2.2.3)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.10/dist-packages (from requests>=2.32.2-
>datasets) (2024.12.14)
Requirement already satisfied: python-dateutil>=2.8.2 in
/usr/local/lib/python3.10/dist-packages (from pandas->datasets)
(2.8.2)
Requirement already satisfied: pytz>=2020.1 in
/usr/local/lib/python3.10/dist-packages (from pandas->datasets)
(2024.2)
Requirement already satisfied: tzdata>=2022.7 in
/usr/local/lib/python3.10/dist-packages (from pandas->datasets)
(2024.2)
Requirement already satisfied: six>=1.5 in
/usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.8.2-
>pandas->datasets) (1.17.0)
Downloading datasets-3.2.0-py3-none-any.whl (480 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 480.6/480.6 kB 16.7 MB/s eta
0:00:00
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 116.3/116.3 kB 8.0 MB/s eta
0:00:00
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 179.3/179.3 kB 13.6 MB/s eta
0:00:00
ultiprocess-0.70.16-py310-none-any.whl (134 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 134.8/134.8 kB 6.8 MB/s eta
0:00:00
anylinux_2_17_x86_64.manylinux2014_x86_64.whl (194 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 194.1/194.1 kB 10.0 MB/s eta
0:00:00
```

```
ultiprocess, datasets
  Attempting uninstall: fsspec
    Found existing installation: fsspec 2024.10.0
    Uninstalling fsspec-2024.10.0:
      Successfully uninstalled fsspec-2024.10.0
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 2024.10.0 requires fsspec==2024.10.0, but you have fsspec
2024.9.0 which is incompatible.
Successfully installed datasets-3.2.0 dill-0.3.8 fsspec-2024.9.0
multiprocess-0.70.16 xxhash-3.5.0
```

```
from datasets import load_dataset
dataset = load_dataset("imdb")
```

```
/usr/local/lib/python3.10/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":"15b4e2280bc04f32870bfeb2fcf7e160","version_major":2,"vers
ion_minor":0}
```

```
{"model_id":"d94eb780c6eb4c569eca30ae4880ddc6","version_major":2,"vers
ion_minor":0}
```

```
{"model_id":"8ecc61bf002945c4a4c5fd4db911a948","version_major":2,"vers
ion_minor":0}
```

```
{"model_id":"064aad20f6974b1ba62f737de7b909b4","version_major":2,"vers
ion_minor":0}
```

```
{"model_id":"1852a39d7b864ae5aab905dacec23703","version_major":2,"vers
ion_minor":0}
```

```
{"model_id":"2e127a905f9d40bba62da8e5cebc489b","version_major":2,"vers
ion_minor":0}
```

```
{"model_id":"efc3a9fb0dc94b479bb51610427b5e41","version_major":2,"vers
ion_minor":0}
```


```
from transformers import AutoTokenizer
tokenizer = AutoTokenizer.from_pretrained("bert-base-uncased")

def tokenize_function(examples):
```

```
    return tokenizer(examples["text"], padding="max_length",
truncation=True)

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

```
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
interrupt this and resume the migration later on by calling
`transformers.utils.move_cache()`.
```


```
{"model_id":"8901dc98707947d6b8322df5c71d2eaf","version_major":2,"vers
ion_minor":0}

{"model_id":"96cbf3869c2a4d399162ed04ffe74756","version_major":2,"vers
ion_minor":0}

{"model_id":"b671a61cdb0940c098a539b4311b01f2","version_major":2,"vers
ion_minor":0}

{"model_id":"a566097c348b41b8817e692dcd7d6b42","version_major":2,"vers
ion_minor":0}

{"model_id":"696125341c564512b88237391e6ad24b","version_major":2,"vers
ion_minor":0}

{"model_id":"08804b0a0c234b70b9b106e8dfb2ef92","version_major":2,"vers
ion_minor":0}

{"model_id":"75bdaec5f23e4cd4a92d476965e12750","version_major":2,"vers
ion_minor":0}

{"model_id":"51295e33232146e6a958484cf6dd2ab3","version_major":2,"vers
ion_minor":0}
```


```
train_test_split =
tokenized_datasets["train"].train_test_split(test_size=0.2)
train_dataset = train_test_split["train"]
eval_dataset = train_test_split["test"]

print(f"Jumlah data pelatihan: {len(train_dataset)}")
print(f"Jumlah data validasi: {len(eval_dataset)}")
```

```
Jumlah data pelatihan: 20000
Jumlah data validasi: 5000
```

```
from torch.utils.data import DataLoader
train_dataloader = DataLoader(train_dataset, shuffle=True,
batch_size=8)
eval_dataloader = DataLoader(eval_dataset, batch_size=8)
```

```
from transformers import AutoModelForSequenceClassification
model = AutoModelForSequenceClassification.from_pretrained("bert-base-uncased", num_labels=2)
```

```
{"model_id":"cdd5d50f3c0249f19092081efbbd1aca","version_major":2,"version_minor":0}
```

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

```
from transformers import TrainingArguments, Trainer

training_args = TrainingArguments(
    output_dir="./results",                # Direktori hasil
    evaluation_strategy="epoch",          # Evaluasi setiap epoch
    learning_rate=2e-5,                   # Learning rate
    per_device_train_batch_size=8,        # Batch size untuk pelatihan
    per_device_eval_batch_size=8,         # Batch size untuk evaluasi
    num_train_epochs=3,                   # Jumlah epoch
    weight_decay=0.01,                    # L2 weight decay
)
trainer = Trainer(
    model=model,
    args=training_args,
    train_dataset=train_dataset,
    eval_dataset=eval_dataset,
)
```

```
/usr/local/lib/python3.10/dist-packages/transformers/training_args.py:1575: FutureWarning: `evaluation_strategy` is deprecated and will be removed in version 4.46 of  Transformers. Use `eval_strategy` instead
  warnings.warn(
```

```
trainer.train()
```

```
<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, or press ctrl+c to quit:wandb: Appending key for api.wandb.ai to your netrc file: /root/.netrc
```

```
<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>

<IPython.core.display.HTML object>

<IPython.core.display.HTML object>

---------------------------------------------------------------------------
KeyboardInterrupt                         Traceback (most recent call last)
<ipython-input-13-3435b262f1ae> in <cell line: 1>()
----> 1 trainer.train()

/usr/local/lib/python3.10/dist-packages/transformers/trainer.py in train(self, resume_from_checkpoint, trial, ignore_keys_for_eval, **kwargs)
   2162                 hf_hub_utils.enable_progress_bars()
   2163         else:
-> 2164             return inner_training_loop(
   2165                 args=args,
   2166                 resume_from_checkpoint=resume_from_checkpoint,

/usr/local/lib/python3.10/dist-packages/transformers/trainer.py in _inner_training_loop(self, batch_size, args, resume_from_checkpoint, trial, ignore_keys_for_eval)
   2527                         args.logging_nan_inf_filter
   2528                         and not is_torch_xla_available()
-> 2529                         and (torch.isnan(tr_loss_step) or torch.isinf(tr_loss_step))
   2530                     ):
   2531                         # if loss is nan or inf simply add the average of previous logged losses

KeyboardInterrupt:
```

```python
import numpy as np
from evaluate import load

metric = load("accuracy")

def compute_metrics(eval_pred):
    logits, labels = eval_pred
    predictions = np.argmax(logits, axis=-1)
    return metric.compute(predictions=predictions, references=labels)
```

```
trainer.evaluate()
```

```
---------------------------------------------------------------------------
-----
ModuleNotFoundError                       Traceback (most recent call
last)
<ipython-input-14-1945aa0b6b4e> in <cell line: 2>()
      1 import numpy as np
----> 2 from evaluate import load
      3
      4 metric = load("accuracy")
      5

ModuleNotFoundError: No module named 'evaluate'

---------------------------------------------------------------------------
-----
NOTE: If your import is failing due to a missing package, you can
manually install dependencies using either !pip or !apt.

To view examples of installing some common dependencies, click the
"Open Examples" button below.
---------------------------------------------------------------------------
-----
```