Text Classification NLP with 10,000 Rotten Tomatoes reviews

I'll be using the Rotten Tomatoes movie review dataset from Hugging Face datasets, this contains 10,000 reviews that will be split for training and testing. In this notebook, we use pretrained sentiment analysis model to classify the movie reviews from either as positive or negative.

Libraries added:

!pip install datasets transformers evaluate

from datasets import load_dataset

from transformers import pipeline, AutoTokenizer, AutoModelForSequenceClassification, TrainingArguments, Trainer

import evaluate

import numpy as np

import torch

from sklearn.metrics import accuracy_score

```
!pip install datasets
!pip install transformers
!pip install evaluate
from datasets import load_dataset
from transformers import pipeline, AutoTokenizer, A
import evaluate
import numpy as np
import torch
from sklearn.metrics import accuracy_score

dataset = load_dataset("rotten_tomatoes")
print(dataset)
dataset['train'][0]
```

ModuleNotFoundError

Please explain the error:

```
ModuleNotFoundError: No m
```

Hypothesis

The error message `ModuleNotFoundError: No module named 'datasets'` indicates that the Python interpreter cannot find the `datasets` module. This is likely because the `datasets` library is not installed in the current environment. The user mentioned installing some libraries in a markdown cell, but that cell would not have been executed as Python code, it is a comment.

Suggested Changes

Install `datasets` library using `pip`:

```
!pip install datasets
```

Use code with caution

And then add:

```
from datasets import load_
```

Use code with caution

to import it.

```
Collecting datasets
  Downloading datasets-3.5.1-py3-none-any.whl.
Requirement already satisfied: filelock in /us
Requirement already satisfied: numpy>=1.17 in
Requirement already satisfied: pyarrow>=15.0.0
Collecting dill<0.3.9,>=0.3.0 (from datasets)
  Downloading dill-0.3.8-py3-none-any.whl.meta
Requirement already satisfied: pandas in /usr/
Requirement already satisfied: requests>=2.32.
Requirement already satisfied: tqdm>=4.66.3 in
Collecting xxhash (from datasets)
  Downloading xxhash-3.5.0-cp311-cp311-manylin
Collecting multiprocess<0.70.17 (from datasets
  Downloading multiprocess-0.70.16-py311-none-
Collecting fsspec<=2025.3.0,>=2023.1.0 (from f
  Downloading fsspec-2025.3.0-py3-none-any.whl
Requirement already satisfied: aiohttp in /usr
Requirement already satisfied: huggingface-hub
Requirement already satisfied: packaging in /u
Requirement already satisfied: pyyaml>=5.1 in
Requirement already satisfied: aiohappyeyeball
Requirement already satisfied: aiosignal>=1.1.
Requirement already satisfied: attrs>=17.3.0 i
Requirement already satisfied: frozenlist>=1.1
Requirement already satisfied: multidict<7.0,>
Requirement already satisfied: propcache>=0.2.
Requirement already satisfied: yarl<2.0,>=1.17
Requirement already satisfied: typing-extensio
Requirement already satisfied: charset-normali
Requirement already satisfied: idna<4,>=2.5 in
Requirement already satisfied: urllib3<3,>=1.2
Requirement already satisfied: certifi>=2017.4
Requirement already satisfied: python-dateutil
Requirement already satisfied: pytz>=2020.1 in
Requirement already satisfied: tzdata>=2022.7
Requirement already satisfied: six>=1.5 in /us
Downloading datasets-3.5.1-py3-none-any.whl (4
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 49
Downloading dill-0.3.8-py3-none-any.whl (116 k
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 11
Downloading fsspec-2025.3.0-py3-none-any.whl (
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 19
Downloading multiprocess-0.70.16-py311-none-an
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 14
Downloading xxhash-3.5.0-cp311-cp311-manylinux
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 19
Installing collected packages: xxhash, fsspec,
  Attempting uninstall: fsspec
    Found existing installation: fsspec 2025.3
    Uninstalling fsspec-2025.3.2:
```

Rate this answer

```
      Successfully uninstalled fsspec-2025.3.2
ERROR: pip's dependency resolver does not curr
torch 2.6.0+cu124 requires nvidia-cublas-cu12=
torch 2.6.0+cu124 requires nvidia-cuda-cupti-c
torch 2.6.0+cu124 requires nvidia-cuda-nvrtc-c
torch 2.6.0+cu124 requires nvidia-cuda-runtime
torch 2.6.0+cu124 requires nvidia-cudnn-cu12==
torch 2.6.0+cu124 requires nvidia-cufft-cu12==
torch 2.6.0+cu124 requires nvidia-curand-cu12=
torch 2.6.0+cu124 requires nvidia-cusolver-cu1
torch 2.6.0+cu124 requires nvidia-cusparse-cu1
torch 2.6.0+cu124 requires nvidia-nvjitlink-cu
gcsfs 2025.3.2 requires fsspec==2025.3.2, but
Successfully installed datasets-3.5.1 dill-0.3
Requirement already satisfied: transformers in
Requirement already satisfied: filelock in /us
Requirement already satisfied: huggingface-hub
Requirement already satisfied: numpy>=1.17 in
Requirement already satisfied: packaging>=20.0
Requirement already satisfied: pyyaml>=5.1 in
Requirement already satisfied: regex!=2019.12.
Requirement already satisfied: requests in /us
Requirement already satisfied: tokenizers<0.22
Requirement already satisfied: safetensors>=0.
Requirement already satisfied: tqdm>=4.27 in /
Requirement already satisfied: fsspec>=2023.5.
Requirement already satisfied: typing-extensio
Requirement already satisfied: charset-normali
Requirement already satisfied: idna<4,>=2.5 in
Requirement already satisfied: urllib3<3,>=1.2
Requirement already satisfied: certifi>=2017.4
Collecting evaluate
  Downloading evaluate-0.4.3-py3-none-any.whl.
Requirement already satisfied: datasets>=2.0.0
Requirement already satisfied: numpy>=1.17 in
Requirement already satisfied: dill in /usr/lo
Requirement already satisfied: pandas in /usr/
Requirement already satisfied: requests>=2.19.
Requirement already satisfied: tqdm>=4.62.1 in
Requirement already satisfied: xxhash in /usr/
Requirement already satisfied: multiprocess in
Requirement already satisfied: fsspec>=2021.05
Requirement already satisfied: huggingface-hub
Requirement already satisfied: packaging in /u
Requirement already satisfied: filelock in /us
Requirement already satisfied: pyarrow>=15.0.0
Requirement already satisfied: aiohttp in /usr
Requirement already satisfied: pyyaml>=5.1 in
Requirement already satisfied: typing-extensio
Requirement already satisfied: charset-normali
```

```
Requirement already satisfied: idna<4,>=2.5 in
Requirement already satisfied: urllib3<3,>=1.2
Requirement already satisfied: certifi>=2017.4
Requirement already satisfied: python-dateutil
Requirement already satisfied: pytz>=2020.1 in
Requirement already satisfied: tzdata>=2022.7
Requirement already satisfied: aiohappyeyeball
Requirement already satisfied: aiosignal>=1.1.
Requirement already satisfied: attrs>=17.3.0 i
Requirement already satisfied: frozenlist>=1.1
Requirement already satisfied: multidict<7.0,>
Requirement already satisfied: propcache>=0.2.
Requirement already satisfied: yarl<2.0,>=1.17
Requirement already satisfied: six>=1.5 in /us
Downloading evaluate-0.4.3-py3-none-any.whl (8
                                              84
Installing collected packages: evaluate
Successfully installed evaluate-0.4.3
/usr/local/lib/python3.11/dist-packages/huggin
The secret `HF_TOKEN` does not exist in your C
To authenticate with the Hugging Face Hub, cre
You will be able to reuse this secret in all o
Please note that authentication is recommended
  warnings.warn(
```

README.md: 100% 7.46k/7.46k [00:00<00:00, 337kB/s]

train.parquet: 100% 699k/699k [00:00<00:00, 5.40MB/s]

validation.parquet: 100% 90.0k/90.0k [00:00<00:00, 4.74MB/s]

test.parquet: 100% 92.2k/92.2k [00:00<00:00, 4.38MB/s]

Generating train split: 100% 8530/8530 [00:00<00:00, 98751.07 e
s]

Generating validation split: 100% 1066/1066 [00:00<00:00, 30803
s]

Generating test split: 100% 1066/1066 [00:00<00:00, 39981.11 ex
s]

```
DatasetDict({
    train: Dataset({
        features: ['text', 'label'],
        num_rows: 8530
    })
    validation: Dataset({
        features: ['text', 'label'],
        num_rows: 1066
    })
```

```
        })
        test: Dataset({
            features: ['text', 'label'],
            num_rows: 1066
        })
    })
```

Model set up and hyperparameter tuning

As I have texts that dont exceed 512 tokens, I chose to continue with the base(distilbert-base-uncased). Alternatively was going to use (allenai/longformer-base-4096), but ran slower than distolbert.

I included my standard tokenizer and classifier onto the dataset that will run preprocessing, passing the inputs through the model, and postprocessing. I had issues with the token limit and adjusted the maximum capable.

My inputs is the preprocess_function that I have defined to also include padding, truncation, and max_length. I also mapped the dataset and batched it to run smoother than defore as its running on a older cpu.

```
checkpoint = "distilbert-base-uncased-finetuned-ss
tokenizer = AutoTokenizer.from_pretrained(checkpo
model = AutoModelForSequenceClassification.from_p
classifier = pipeline("sentiment-analysis", model=

# Input to preprocess the dataset with the define
def preprocess_function(examples):
    return tokenizer(examples["text"], padding="ma

dataset = dataset.map(preprocess_function, batched
```

```
tokenizer_config.json: 100%  48.0/48.0 [00:00<00:00, 2.45kB/s]
config.json: 100%  629/629 [00:00<00:00, 21.7kB/s]
vocab.txt: 100%  232k/232k [00:00<00:00, 4.68MB/s]
Xet Storage is enabled for this repo, but the
WARNING:huggingface_hub.file_download:Xet Stor
model.safetensors: 100%  268M/268M [00:01<00:00, 190MB/s]
Device set to use cpu
Map: 100%  8530/8530 [00:08<00:00, 1088.03 examples/
s]
Map: 100%  1066/1066 [00:01<00:00, 1104.04 examples/
s]
```

Testing

To make sure that the classifer is working, I test it with the first 50 rows and make the display its predictions. Interestingly enough, this model needs defined words especially for a label_map of the results. Label_map works just fine without it, but kept it in there for transparency.

```
texts = dataset["test"]["text"][:50]
results = classifier(texts)
label_map = {"NEGATIVE": "Negative", "POSITIVE": '
```

```
for text, result in zip(texts, results):
    print(f"Review: {text}")
    print(f"Label: {label_map[result['label']]}, S
    print("-" * 80) # This made it easier to read
```

```
--------------------------------------------------
Review: scherfig's light-hearted profile of em
Label: Positive, Score: 0.9999
--------------------------------------------------
Review: a journey spanning nearly three decade
Label: Positive, Score: 0.9996
--------------------------------------------------
Review: the wonderfully lush morvern callar is
Label: Positive, Score: 0.9994
--------------------------------------------------
Review: as it turns out , you can go home agai
Label: Positive, Score: 0.9997
--------------------------------------------------
Review: you've already seen city by the sea ur
Label: Positive, Score: 0.9981
--------------------------------------------------
Review: this kind of hands-on storytelling is
Label: Positive, Score: 0.9997
--------------------------------------------------
Review: making such a tragedy the backdrop to
Label: Positive, Score: 0.9992
--------------------------------------------------
Review: grown-up quibbles are beside the point
Label: Positive, Score: 0.9983
--------------------------------------------------
Review: a powerful , chilling , and affecting
Label: Positive, Score: 0.9998
--------------------------------------------------
Review: this is a fascinating film because the
Label: Positive, Score: 0.9997
--------------------------------------------------
Review: a dreadful day in irish history is giv
Label: Positive, Score: 0.9980
--------------------------------------------------
Review: . . . a good film that must have baffl
Label: Positive, Score: 0.9998
--------------------------------------------------
Review: . . . is funny in the way that makes y
Label: Positive, Score: 0.9999
--------------------------------------------------
Review: devotees of star trek ii : the wrath o
Label: Negative, Score: 0.9994
--------------------------------------------------
Review: a soul-stirring documentary about the
```

```
Label: Positive, Score: 0.9999
--------------------------------------------------
Review: what's so striking about jolie's perfo
Label: Positive, Score: 0.9998
--------------------------------------------------
Review: the main story . . . is compelling enc
Label: Negative, Score: 0.9994
--------------------------------------------------
Review: the performances are immaculate , with
Label: Positive, Score: 0.9999
--------------------------------------------------
Review: kinnear . . . gives his best screen pe
Label: Positive, Score: 0.9998
--------------------------------------------------
Review: hugh grant , who has a good line in ch
Label: Positive, Score: 0.9996
```

Evaluate accuracy on the full dataset

Evaluating the accuracy on all test reviews and make my own predictions. I then convert the labels into integers to make the accuracy. I can't use the BLEU method as that is focused for other metrics especially for summarizing and translation. As I am using sentiment analysis, we'll use accuracy.

```
accuracy = evaluate.load("accuracy")
test_texts = dataset["test"]["text"]
test_labels = dataset["test"]["label"]
predictions = classifier(test_texts)

predicted_labels = [1 if result["label"] == "POSI
accuracy_result = accuracy.compute(predictions=pre
print("Test Set Accuracy:", accuracy_result["accu
```

```
Downloading builder script: 100%  4.20k/4.20k [00:00<00:00, 74.7
Test Set Accuracy: 0.8968105065666041
```

Final Reflection

- **Model Used**: distilbert-base-uncased-finetuned-sst-2-english
- **Dataset**: Rotten Tomatoes movie review dataset (binary labels)
- **Process**: Tokenize → Predict → Evaluate

Key Learnings

- Useful in this case for evaluating the success/failure of a film to set proprt ratings.
- Using pretrained models allows for quick and effective sentiment classification.
- The Hugging Face pipeline makes inference simple and interpretable.
- Accuracy on test data is strong with 89%
- Would need new measure of metrics if attempting summarization or translation.

Limitations

- Doesn't pull any key words or phrases commonly used.
- The model only supports binary classification.
- It may miss contextual cues not seen during training.

Future Improvements

- Explore multi-class sentiment classification.
- Add model interpretability.
- Try more advanced models like RoBERTa or XLNet.
- Try other datasets with more data like with "IMDB" with 50,000 reviews as initially attempted.

Performing the sentiment analysis

I apply the classifier just as an option and was able to get the full list below. Took around 11 minutes to complete. Included the tokens as the output for the training of the dataset.

```
preds = classifier(dataset['train']['text'])
preds
```

```
 {'label': 'POSITIVE', 'score':
0.997366726398468},
 {'label': 'POSITIVE', 'score':
0.9983170032501221},
 {'label': 'POSITIVE', 'score':
0.9996293783187866},
 {'label': 'POSITIVE', 'score':
0.9998421669006348},
 {'label': 'POSITIVE', 'score':
0.9986334443092346},
 {'label': 'POSITIVE', 'score':
0.9998315572738647},
 {'label': 'POSITIVE', 'score':
0.9997726082801819},
 {'label': 'POSITIVE', 'score':
0.9998014569282532},
 {'label': 'POSITIVE', 'score':
0.9992374181747437},
 {'label': 'POSITIVE', 'score':
0.9970617890357971},
 {'label': 'POSITIVE', 'score':
0.998033344745636},
 {'label': 'POSITIVE', 'score':
0.9997692704200745},
 {'label': 'POSITIVE', 'score':
0.9996053576469421},
 {'label': 'NEGATIVE', 'score':
0.9979937076568604},
 {'label': 'POSITIVE', 'score':
0.9804105162620544},
 {'label': 'POSITIVE', 'score':
0.9998407363891602},
 {'label': 'POSITIVE', 'score':
```

```
0.9959083795547485},
 {'label': 'POSITIVE', 'score':
0.999883770942688},
 {'label': 'POSITIVE', 'score':
0.9994814991950989},
 {'label': 'POSITIVE', 'score':
0.9587122201919556},
 {'label': 'POSITIVE', 'score':
0.9998636245727539},
 {'label': 'POSITIVE', 'score':
0.9985867738723755},
 {'label': 'POSITIVE', 'score':
0.9998760223388672},
 {'label': 'POSITIVE', 'score':
0.9993273019790649},
 {'label': 'POSITIVE', 'score':
0.9972118735313416},
 {'label': 'POSITIVE', 'score':
0.9953987002372742},
 {'label': 'POSITIVE', 'score':
0.9974973797798157},
 {'label': 'POSITIVE', 'score':
0.9974199533462524},
 {'label': 'POSITIVE', 'score':
0.9990969896316528},
 {'label': 'POSITIVE', 'score':
0.9966320395469666}.
```

Start coding or generate with AI.

Enter a prompt here

0 / 2000

Gemini can make mistakes so double-check responses and use code with caution. Learn more