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, AutoModelForSequenceClassificatic
import evaluate
import numpy as np
import torch
from sklearn.metrics import accuracy_score

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

```
Collecting datasets
  Downloading datasets-3.5.1-py3-none-any.whl.metadata (19 kB)
```

```
Requirement already satisfied: filelock in /usr/local/lib/python3.11/dist-pack
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.11/dist-p
Requirement already satisfied: pyarrow>=15.0.0 in /usr/local/lib/python3.11/di
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-packag
Requirement already satisfied: requests>=2.32.2 in /usr/local/lib/python3.11/d
Requirement already satisfied: tqdm>=4.66.3 in /usr/local/lib/python3.11/dist-
Collecting xxhash (from datasets)
  Downloading xxhash-3.5.0-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86
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
  Downloading fsspec-2025.3.0-py3-none-any.whl.metadata (11 kB)
Requirement already satisfied: aiohttp in /usr/local/lib/python3.11/dist-packa
Requirement already satisfied: huggingface-hub>=0.24.0 in /usr/local/lib/pytho
Requirement already satisfied: packaging in /usr/local/lib/python3.11/dist-pac
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.11/dist-p
Requirement already satisfied: aiohappyeyeballs>=2.3.0 in /usr/local/lib/pytho
Requirement already satisfied: aiosignal>=1.1.2 in /usr/local/lib/python3.11/d
Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.11/dist
Requirement already satisfied: frozenlist>=1.1.1 in /usr/local/lib/python3.11/
Requirement already satisfied: multidict<7.0,>=4.5 in /usr/local/lib/python3.1
Requirement already satisfied: propcache>=0.2.0 in /usr/local/lib/python3.11/d
Requirement already satisfied: yarl<2.0,>=1.17.0 in /usr/local/lib/python3.11/
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/py
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/pyth
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.11
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.11/dist-
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.11/dis
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.11/dist-pack
Downloading datasets-3.5.1-py3-none-any.whl (491 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 491.4/491.4 kB 24.3 MB/s eta 0:00:
Downloading dill-0.3.8-py3-none-any.whl (116 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 116.3/116.3 kB 8.8 MB/s eta 0:00:0
Downloading fsspec-2025.3.0-py3-none-any.whl (193 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 193.6/193.6 kB 12.2 MB/s eta 0:00:
Downloading multiprocess-0.70.16-py311-none-any.whl (143 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 143.5/143.5 kB 9.9 MB/s eta 0:00:0
Downloading xxhash-3.5.0-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_6
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 194.8/194.8 kB 16.0 MB/s eta 0:00:
Installing collected packages: xxhash, fsspec, dill, multiprocess, datasets
  Attempting uninstall: fsspec
    Found existing installation: fsspec 2025.3.2
    Uninstalling fsspec-2025.3.2:
      Successfully uninstalled fsspec-2025.3.2
ERROR: pip's dependency resolver does not currently take into account all the
torch 2.6.0+cu124 requires nvidia-cublas-cu12==12.4.5.8; platform_system == "I
```

```
torch 2.6.0+cu124 requires nvidia-cuda-cupti-cu12==12.4.127; platform_system =
torch 2.6.0+cu124 requires nvidia-cuda-nvrtc-cu12==12.4.127; platform_system =
torch 2.6.0+cu124 requires nvidia-cuda-runtime-cu12==12.4.127; platform_system
torch 2.6.0+cu124 requires nvidia-cudnn-cu12==9.1.0.70; platform_system == "Li
torch 2.6.0+cu124 requires nvidia-cufft-cu12==11.2.1.3; platform_system == "Li
torch 2.6.0+cu124 requires nvidia-curand-cu12==10.3.5.147; platform_system ==
torch 2.6.0+cu124 requires nvidia-cusolver-cu12==11.6.1.9; platform_system ==
torch 2.6.0+cu124 requires nvidia-cusparse-cu12==12.3.1.170; platform_system =
torch 2.6.0+cu124 requires nvidia-nvjitlink-cu12==12.4.127; platform_system ==
gcsfs 2025.3.2 requires fsspec==2025.3.2, but you have fsspec 2025.3.0 which i
Successfully installed datasets-3.5.1 dill-0.3.8 fsspec-2025.3.0 multiprocess-
Requirement already satisfied: transformers in /usr/local/lib/python3.11/dist-
Requirement already satisfied: filelock in /usr/local/lib/python3.11/dist-pack
Requirement already satisfied: huggingface-hub<1.0,>=0.30.0 in /usr/local/lib/
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.11/dist-p
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.11/di
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.11/dist-p
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.11/
Requirement already satisfied: requests in /usr/local/lib/python3.11/dist-pack
Requirement already satisfied: tokenizers<0.22,>=0.21 in /usr/local/lib/python
Requirement already satisfied: safetensors>=0.4.3 in /usr/local/lib/python3.11
Requirement already satisfied: tqdm>=4.27 in /usr/local/lib/python3.11/dist-pa
Requirement already satisfied: fsspec>=2023.5.0 in /usr/local/lib/python3.11/d
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/py
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/pyth
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.11
Collecting evaluate
  Downloading evaluate-0.4.3-py3-none-any.whl.metadata (9.2 kB)
Requirement already satisfied: datasets>=2.0.0 in /usr/local/lib/python3.11/di
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.11/dist-p
Requirement already satisfied: dill in /usr/local/lib/python3.11/dist-packages
Requirement already satisfied: pandas in /usr/local/lib/python3.11/dist-packag
Requirement already satisfied: requests>=2.19.0 in /usr/local/lib/python3.11/d
Requirement already satisfied: tqdm>=4.62.1 in /usr/local/lib/python3.11/dist-
Requirement already satisfied: xxhash in /usr/local/lib/python3.11/dist-packag
Requirement already satisfied: multiprocess in /usr/local/lib/python3.11/dist-
Requirement already satisfied: fsspec>=2021.05.0 in /usr/local/lib/python3.11/
Requirement already satisfied: huggingface-hub>=0.7.0 in /usr/local/lib/python
Requirement already satisfied: packaging in /usr/local/lib/python3.11/dist-pac
Requirement already satisfied: filelock in /usr/local/lib/python3.11/dist-pack
Requirement already satisfied: pyarrow>=15.0.0 in /usr/local/lib/python3.11/di
Requirement already satisfied: aiohttp in /usr/local/lib/python3.11/dist-packa
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.11/dist-p
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/py
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/pyth
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.11
```

```
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.11/dist-
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.11/dis
Requirement already satisfied: aiohappyeyeballs>=2.3.0 in /usr/local/lib/pytho
Requirement already satisfied: aiosignal>=1.1.2 in /usr/local/lib/python3.11/d
Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.11/dist
Requirement already satisfied: frozenlist>=1.1.1 in /usr/local/lib/python3.11/
Requirement already satisfied: multidict<7.0,>=4.5 in /usr/local/lib/python3.1
Requirement already satisfied: propcache>=0.2.0 in /usr/local/lib/python3.11/d
Requirement already satisfied: yarl<2.0,>=1.17.0 in /usr/local/lib/python3.11/
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.11/dist-pack
Downloading evaluate-0.4.3-py3-none-any.whl (84 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 84.0/84.0 kB 7.2 MB/s eta 0:00:00
Installing collected packages: evaluate
Successfully installed evaluate-0.4.3
/usr/local/lib/python3.11/dist-packages/huggingface_hub/utils/_auth.py:94: Use
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
You will be able to reuse this secret in all of your notebooks.
Please note that authentication is recommended but still optional to access pu
  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 examples/s]

Generating validation split: 100% 1066/1066 [00:00<00:00, 30803.29 examples/s]

Generating test split: 100% 1066/1066 [00:00<00:00, 39981.11 examples/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
```

```
            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-sst-2-english"
tokenizer = AutoTokenizer.from_pretrained(checkpoint)
model = AutoModelForSequenceClassification.from_pretrained(checkpoint)
classifier = pipeline("sentiment-analysis", model=model, tokenizer=tokenizer)

# Input to preprocess the dataset with the defined padding, truncation and max_le
def preprocess_function(examples):
    return tokenizer(examples["text"], padding="max_length", truncation=True, max_

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

```
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 'hf_xet' package is not installe
WARNING:huggingface_hub.file_download:Xet Storage is enabled for this repo, bu
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]
Map: 100%  1066/1066 [00:01<00:00, 553.64 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": "Positive"}

for text, result in zip(texts, results):
    print(f"Review: {text}")
    print(f"Label: {label_map[result['label']]}, Score: {result['score']:.4f}")
    print("-" * 80) # This made it easier to read
```

```
print('-' * 80) # this made it easier to read
```

```
view: what's so striking about jolie's performance is that she never lets her
bel: Positive, Score: 0.9998
--------------------------------------------------------------------------------
view: the main story . . . is compelling enough , but it's difficult to shrug
bel: Negative, Score: 0.9994
--------------------------------------------------------------------------------
view: the performances are immaculate , with roussillon providing comic relief
bel: Positive, Score: 0.9999
--------------------------------------------------------------------------------
view: kinnear . . . gives his best screen performance with an oddly winning po
bel: Positive, Score: 0.9998
--------------------------------------------------------------------------------
view: hugh grant , who has a good line in charm , has never been more charming
bel: Positive, Score: 0.9996
--------------------------------------------------------------------------------
view: there's a lot of tooth in roger dodger . but what's nice is that there's
bel: Positive, Score: 0.9997
--------------------------------------------------------------------------------
view: reminiscent of alfred hitchcock's thrillers , most of the scary parts in
bel: Positive, Score: 0.9932
--------------------------------------------------------------------------------
view: one of the best looking and stylish animated movies in quite a while . .
bel: Positive, Score: 0.9999
--------------------------------------------------------------------------------
view: its use of the thriller form to examine the labyrinthine ways in which p
bel: Positive, Score: 0.9998
--------------------------------------------------------------------------------
view: denver should not get the first and last look at one of the most triumph
bel: Positive, Score: 0.9144
--------------------------------------------------------------------------------
view: you needn't be steeped in '50s sociology , pop culture or movie lore to
bel: Positive, Score: 0.9903
--------------------------------------------------------------------------------
view: waiting for godard can be fruitful : 'in praise of love' is the director
bel: Positive, Score: 0.9996
--------------------------------------------------------------------------------
view: a gangster movie with the capacity to surprise .
bel: Positive, Score: 0.9998
--------------------------------------------------------------------------------
view: the film has a laundry list of minor shortcomings , but the numerous sce
bel: Positive, Score: 0.9995
--------------------------------------------------------------------------------
view: if not a home run , then at least a solid base hit .
bel: Positive, Score: 0.9997
--------------------------------------------------------------------------------
view: goldmember is funny enough to justify the embarrassment of bringing a ba
bel: Positive, Score: 0.9943
--------------------------------------------------------------------------------
```

```
view: . . . a fairly disposable yet still entertaining b picture .
bel: Positive, Score: 0.9973
--------------------------------------------------------------------------------
view: it may not be particularly innovative , but the film's crisp , unaffecte
bel: Positive, Score: 0.9999
--------------------------------------------------------------------------------
view: the film truly does rescue [the funk brothers] from motown's shadows . i
bel: Positive, Score: 0.9996
--------------------------------------------------------------------------------
view: drawing on an irresistible , languid romanticism , byler reveals the way
bel: Positive, Score: 0.9998
--------------------------------------------------------------------------------
```

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"] == "POSITIVE" else 0 for result in predict
accuracy_result = accuracy.compute(predictions=predicted_labels, references=test_la
print("Test Set Accuracy:", accuracy_result["accuracy"])
```

```
Downloading builder script: 100%  4.20k/4.20k [00:00<00:00, 74.7kB/s]
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': 0.9998360872268677},
 {'label': 'POSITIVE', 'score': 0.9998277425765991},
 {'label': 'NEGATIVE', 'score': 0.9960036873817444},
 {'label': 'POSITIVE', 'score': 0.9998257756233215},
 {'label': 'POSITIVE', 'score': 0.9997782111167908},
 {'label': 'POSITIVE', 'score': 0.9998192191123962},
 {'label': 'POSITIVE', 'score': 0.9998753070831299},
 {'label': 'NEGATIVE', 'score': 0.9845964312553406},
 {'label': 'POSITIVE', 'score': 0.997896671295166},
 {'label': 'POSITIVE', 'score': 0.9998527765274048},
 {'label': 'POSITIVE', 'score': 0.9998834133148193},
 {'label': 'POSITIVE', 'score': 0.9998408555984497},
 {'label': 'POSITIVE', 'score': 0.9985002279281616},
 {'label': 'POSITIVE', 'score': 0.9998641014099121},
 {'label': 'POSITIVE', 'score': 0.9717578291893005},
 {'label': 'POSITIVE', 'score': 0.9998157620429993},
 {'label': 'POSITIVE', 'score': 0.9992558360099792},
 {'label': 'POSITIVE', 'score': 0.9998703002929688},
 {'label': 'POSITIVE', 'score': 0.9996273517608643},
 {'label': 'POSITIVE', 'score': 0.9998793601989746},
 {'label': 'POSITIVE', 'score': 0.999810516834259},
 {'label': 'POSITIVE', 'score': 0.999845027923584},
 {'label': 'POSITIVE', 'score': 0.9997264742851257},
 {'label': 'POSITIVE', 'score': 0.9979863166809082},
 {'label': 'POSITIVE', 'score': 0.9998657703399658},
 {'label': 'POSITIVE', 'score': 0.9998432397842407},
 {'label': 'POSITIVE', 'score': 0.999482274055481},
 {'label': 'POSITIVE', 'score': 0.9998669624328613},
 {'label': 'NEGATIVE', 'score': 0.8285810351371765},
 {'label': 'POSITIVE', 'score': 0.9589256644248962},
 {'label': 'POSITIVE', 'score': 0.9994179010391235},
 {'label': 'POSITIVE', 'score': 0.9997997879981995},
 {'label': 'POSITIVE', 'score': 0.9888836741447449},
 {'label': 'NEGATIVE', 'score': 0.9983893632888794},
 {'label': 'POSITIVE', 'score': 0.9997174143791199},
 {'label': 'NEGATIVE', 'score': 0.9821356534957886},
 {'label': 'POSITIVE', 'score': 0.999777615070343},
 {'label': 'POSITIVE', 'score': 0.9693498015403748},
 {'label': 'POSITIVE', 'score': 0.999808132648468},
 {'label': 'NEGATIVE', 'score': 0.993931770324707},
 {'label': 'POSITIVE', 'score': 0.9995630383491516},
 {'label': 'POSITIVE', 'score': 0.8148062825202942},
 {'label': 'POSITIVE', 'score': 0.9995459914207458},
 {'label': 'NEGATIVE', 'score': 0.775640070438385},
 {'label': 'POSITIVE', 'score': 0.9997749924659729},
 {'label': 'POSITIVE', 'score': 0.9945492148399353},
 {'label': 'POSITIVE', 'score': 0.999873280525207 5},
 {'label': 'POSITIVE', 'score': 0.999721109867096},
```

```
{'label': 'POSITIVE', 'score': 0.9998144507408142},
{'label': 'POSITIVE', 'score': 0.9979762434959412},
{'label': 'POSITIVE', 'score': 0.9998742341995239},
{'label': 'POSITIVE', 'score': 0.9980592131614685},
{'label': 'POSITIVE', 'score': 0.9995843768119812},
{'label': 'POSITIVE', 'score': 0.9997560381889343},
{'label': 'POSITIVE', 'score': 0.9991065859794617},
{'label': 'POSITIVE', 'score': 0.9973809123039246},
{'label': 'POSITIVE', 'score': 0.9993957281112671},
{'label': 'POSITIVE', 'score': 0.9992929697036743},
{'label': 'POSITIVE', 'score': 0.9998443126678467},
```

Start coding or generate with AI.