Lab6-Assignment: Topic Classification

Use the same training, development, and test partitions of the the 20 newsgroups text dataset as in Lab6.4-Topic-classification-BERT.ipynb

- Fine-tune and examine the performance of another transformer-based pretrained language models, e.g., RoBERTa, XLNet

- Compare the performance of this model to the results achieved in Lab6.4-Topic-classification-BERT.ipynb and to a conventional machine learning approach (e.g., SVM, Naive Bayes) using bag-of-words or other engineered features of your choice. Describe the differences in performance in terms of Precision, Recall, and F1-score evaluation metrics.

```
!pip install simpletransformers
```

```
Looking in indexes: https://pypi.org/simple, https://us-
python.pkg.dev/colab-wheels/public/simple/
Collecting simpletransformers
  Downloading simpletransformers-0.63.9-py3-none-any.whl (250 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 250.5/250.5 KB 4.6 MB/s eta
0:00:00
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 43.6/43.6 KB 2.3 MB/s eta
0:00:00
etadata (setup.py) ... ent already satisfied: tqdm>=4.47.0 in
/usr/local/lib/python3.9/dist-packages (from simpletransformers)
(4.65.0)
Collecting streamlit
  Downloading streamlit-1.20.0-py2.py3-none-any.whl (9.6 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 9.6/9.6 MB 29.2 MB/s eta
0:00:00
anylinux_2_17_x86_64.manylinux2014_x86_64.whl (7.6 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 7.6/7.6 MB 26.0 MB/s eta
0:00:00
ent already satisfied: tensorboard in /usr/local/lib/python3.9/dist-
packages (from simpletransformers) (2.11.2)
Requirement already satisfied: scikit-learn in
/usr/local/lib/python3.9/dist-packages (from simpletransformers)
(1.2.2)
Collecting sentencepiece
  Downloading sentencepiece-0.1.97-cp39-cp39-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl (1.3 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 1.3/1.3 MB 27.5 MB/s eta
0:00:00
ers>=4.6.0
  Downloading transformers-4.26.1-py3-none-any.whl (6.3 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 6.3/6.3 MB 37.1 MB/s eta
0:00:00
ent already satisfied: regex in /usr/local/lib/python3.9/dist-packages
```

```
(from simpletransformers) (2022.6.2)
Requirement already satisfied: numpy in /usr/local/lib/python3.9/dist-
packages (from simpletransformers) (1.22.4)
Collecting datasets
  Downloading datasets-2.10.1-py3-none-any.whl (469 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 469.0/469.0 KB 35.5 MB/s eta
0:00:00
ent already satisfied: scipy in /usr/local/lib/python3.9/dist-packages
(from simpletransformers) (1.10.1)
Requirement already satisfied: pandas in
/usr/local/lib/python3.9/dist-packages (from simpletransformers)
(1.4.4)
Collecting wandb>=0.10.32
  Downloading wandb-0.13.11-py3-none-any.whl (2.0 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 2.0/2.0 MB 27.3 MB/s eta
0:00:00
ent already satisfied: requests in /usr/local/lib/python3.9/dist-
packages (from simpletransformers) (2.25.1)
Requirement already satisfied: pyyaml>=5.1 in
/usr/local/lib/python3.9/dist-packages (from transformers>=4.6.0-
>simpletransformers) (6.0)
Requirement already satisfied: filelock in
/usr/local/lib/python3.9/dist-packages (from transformers>=4.6.0-
>simpletransformers) (3.9.0)
Collecting huggingface-hub<1.0,>=0.11.0
  Downloading huggingface_hub-0.13.2-py3-none-any.whl (199 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 199.2/199.2 KB 30.8 MB/s eta
0:00:00
ent already satisfied: packaging>=20.0 in
/usr/local/lib/python3.9/dist-packages (from transformers>=4.6.0-
>simpletransformers) (23.0)
Requirement already satisfied: psutil>=5.0.0 in
/usr/local/lib/python3.9/dist-packages (from wandb>=0.10.32-
>simpletransformers) (5.4.8)
Collecting sentry-sdk>=1.0.0
  Downloading sentry_sdk-1.16.0-py2.py3-none-any.whl (184 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 184.3/184.3 KB 27.0 MB/s eta
0:00:00
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 184.3/184.3 KB 27.1 MB/s eta
0:00:00
etadata (setup.py) ... ent already satisfied: setuptools in
/usr/local/lib/python3.9/dist-packages (from wandb>=0.10.32-
>simpletransformers) (63.4.3)
Requirement already satisfied: typing-extensions in
/usr/local/lib/python3.9/dist-packages (from wandb>=0.10.32-
>simpletransformers) (4.5.0)
Collecting appdirs>=1.4.3
  Downloading appdirs-1.4.4-py2.py3-none-any.whl (9.6 kB)
Requirement already satisfied: protobuf!=4.21.0,<5,>=3.15.0 in
/usr/local/lib/python3.9/dist-packages (from wandb>=0.10.32-
>simpletransformers) (3.19.6)
Requirement already satisfied: Click!=8.0.0,>=7.0 in
/usr/local/lib/python3.9/dist-packages (from wandb>=0.10.32-
>simpletransformers) (8.1.3)
Collecting setproctitle
  Downloading setproctitle-1.3.2-cp39-cp39-
manylinux_2_5_x86_64.manylinux1_x86_64.manylinux_2_17_x86_64.manylinux
2014_x86_64.whl (30 kB)
Collecting docker-pycreds>=0.4.0
  Downloading docker_pycreds-0.4.0-py2.py3-none-any.whl (9.0 kB)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.9/dist-packages (from requests-
>simpletransformers) (2022.12.7)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in
/usr/local/lib/python3.9/dist-packages (from requests-
>simpletransformers) (1.26.15)
Requirement already satisfied: chardet<5,>=3.0.2 in
/usr/local/lib/python3.9/dist-packages (from requests-
>simpletransformers) (4.0.0)
Requirement already satisfied: idna<3,>=2.5 in
/usr/local/lib/python3.9/dist-packages (from requests-
>simpletransformers) (2.10)
Collecting dill<0.3.7,>=0.3.0
  Downloading dill-0.3.6-py3-none-any.whl (110 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 110.5/110.5 KB 17.8 MB/s eta
0:00:00
anylinux_2_17_x86_64.manylinux2014_x86_64.whl (1.0 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 1.0/1.0 MB 78.4 MB/s eta
0:00:00
ent already satisfied: pyarrow>=6.0.0 in
/usr/local/lib/python3.9/dist-packages (from datasets-
>simpletransformers) (9.0.0)
Requirement already satisfied: fsspec[http]>=2021.11.1 in
/usr/local/lib/python3.9/dist-packages (from datasets-
>simpletransformers) (2023.3.0)
Collecting xxhash
  Downloading xxhash-3.2.0-cp39-cp39-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl (212 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 212.2/212.2 KB 27.7 MB/s eta
0:00:00
ultiprocess
  Downloading multiprocess-0.70.14-py39-none-any.whl (132 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 132.9/132.9 KB 22.1 MB/s eta
0:00:00
ent already satisfied: python-dateutil>=2.8.1 in
/usr/local/lib/python3.9/dist-packages (from pandas-
>simpletransformers) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in
/usr/local/lib/python3.9/dist-packages (from pandas-
>simpletransformers) (2022.7.1)
```

```
Requirement already satisfied: threadpoolctl>=2.0.0 in
/usr/local/lib/python3.9/dist-packages (from scikit-learn-
>simpletransformers) (3.1.0)
Requirement already satisfied: joblib>=1.1.1 in
/usr/local/lib/python3.9/dist-packages (from scikit-learn-
>simpletransformers) (1.1.1)
Requirement already satisfied: importlib-metadata>=1.4 in
/usr/local/lib/python3.9/dist-packages (from streamlit-
>simpletransformers) (6.0.0)
Requirement already satisfied: cachetools>=4.0 in
/usr/local/lib/python3.9/dist-packages (from streamlit-
>simpletransformers) (5.3.0)
Requirement already satisfied: toml in /usr/local/lib/python3.9/dist-
packages (from streamlit->simpletransformers) (0.10.2)
Requirement already satisfied: pillow>=6.2.0 in
/usr/local/lib/python3.9/dist-packages (from streamlit-
>simpletransformers) (8.4.0)
Collecting pydeck>=0.1.dev5
  Downloading pydeck-0.8.0-py2.py3-none-any.whl (4.7 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 4.7/4.7 MB 106.8 MB/s eta
0:00:00
ent already satisfied: tzlocal>=1.1 in /usr/local/lib/python3.9/dist-
packages (from streamlit->simpletransformers) (4.2)
Collecting validators>=0.2
  Downloading validators-0.20.0.tar.gz (30 kB)
  Preparing metadata (setup.py) ... ver
  Downloading semver-2.13.0-py2.py3-none-any.whl (12 kB)
Collecting blinker>=1.0.0
  Downloading blinker-1.5-py2.py3-none-any.whl (12 kB)
Requirement already satisfied: tornado>=6.0.3 in
/usr/local/lib/python3.9/dist-packages (from streamlit-
>simpletransformers) (6.2)
Collecting watchdog
  Downloading watchdog-2.3.1-py3-none-manylinux2014_x86_64.whl (80 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 80.6/80.6 KB 12.7 MB/s eta
0:00:00
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 238.7/238.7 KB 34.0 MB/s eta
0:00:00
ent already satisfied: altair<5,>=3.2.0 in
/usr/local/lib/python3.9/dist-packages (from streamlit-
>simpletransformers) (4.2.2)
Collecting pympler>=0.9
  Downloading Pympler-1.0.1-py3-none-any.whl (164 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 164.8/164.8 KB 24.9 MB/s eta
0:00:00
ent already satisfied: tensorboard-data-server<0.7.0,>=0.6.0 in
/usr/local/lib/python3.9/dist-packages (from tensorboard-
>simpletransformers) (0.6.1)
Requirement already satisfied: tensorboard-plugin-wit>=1.6.0 in
/usr/local/lib/python3.9/dist-packages (from tensorboard-
>simpletransformers) (1.8.1)
Requirement already satisfied: absl-py>=0.4 in
/usr/local/lib/python3.9/dist-packages (from tensorboard-
>simpletransformers) (1.4.0)
Requirement already satisfied: google-auth-oauthlib<0.5,>=0.4.1 in
/usr/local/lib/python3.9/dist-packages (from tensorboard-
>simpletransformers) (0.4.6)
Requirement already satisfied: grpcio>=1.24.3 in
/usr/local/lib/python3.9/dist-packages (from tensorboard-
>simpletransformers) (1.51.3)
Requirement already satisfied: wheel>=0.26 in
/usr/local/lib/python3.9/dist-packages (from tensorboard-
>simpletransformers) (0.38.4)
Requirement already satisfied: markdown>=2.6.8 in
/usr/local/lib/python3.9/dist-packages (from tensorboard-
>simpletransformers) (3.4.1)
Requirement already satisfied: google-auth<3,>=1.6.3 in
/usr/local/lib/python3.9/dist-packages (from tensorboard-
>simpletransformers) (2.16.2)
Requirement already satisfied: werkzeug>=1.0.1 in
/usr/local/lib/python3.9/dist-packages (from tensorboard-
>simpletransformers) (2.2.3)
Requirement already satisfied: entrypoints in
/usr/local/lib/python3.9/dist-packages (from altair<5,>=3.2.0-
>streamlit->simpletransformers) (0.4)
Requirement already satisfied: toolz in /usr/local/lib/python3.9/dist-
packages (from altair<5,>=3.2.0->streamlit->simpletransformers)
(0.12.0)
Requirement already satisfied: jinja2 in
/usr/local/lib/python3.9/dist-packages (from altair<5,>=3.2.0-
>streamlit->simpletransformers) (3.1.2)
Requirement already satisfied: jsonschema>=3.0 in
/usr/local/lib/python3.9/dist-packages (from altair<5,>=3.2.0-
>streamlit->simpletransformers) (4.3.3)
Requirement already satisfied: six>=1.4.0 in
/usr/local/lib/python3.9/dist-packages (from docker-pycreds>=0.4.0-
>wandb>=0.10.32->simpletransformers) (1.15.0)
Requirement already satisfied: attrs>=17.3.0 in
/usr/local/lib/python3.9/dist-packages (from aiohttp->datasets-
>simpletransformers) (22.2.0)
Collecting frozenlist>=1.1.1
  Downloading frozenlist-1.3.3-cp39-cp39-
manylinux_2_5_x86_64.manylinux1_x86_64.manylinux_2_17_x86_64.manylinux
2014_x86_64.whl (158 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 158.8/158.8 KB 23.4 MB/s eta
0:00:00
anylinux_2_17_x86_64.manylinux2014_x86_64.whl (264 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 264.6/264.6 KB 33.2 MB/s eta
0:00:00
alizer<4.0,>=2.0
```

```
  Downloading charset_normalizer-3.1.0-cp39-cp39-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl (199 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 199.2/199.2 KB 20.0 MB/s eta
0:00:00
eout<5.0,>=4.0.0a3
  Downloading async_timeout-4.0.2-py3-none-any.whl (5.8 kB)
Collecting multidict<7.0,>=4.5
  Downloading multidict-6.0.4-cp39-cp39-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl (114 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 114.2/114.2 KB 16.5 MB/s eta
0:00:00
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 62.7/62.7 KB 9.5 MB/s eta
0:00:00
ent already satisfied: rsa<5,>=3.1.4 in /usr/local/lib/python3.9/dist-
packages (from google-auth<3,>=1.6.3->tensorboard->simpletransformers)
(4.9)
Requirement already satisfied: pyasn1-modules>=0.2.1 in
/usr/local/lib/python3.9/dist-packages (from google-auth<3,>=1.6.3-
>tensorboard->simpletransformers) (0.2.8)
Requirement already satisfied: requests-oauthlib>=0.7.0 in
/usr/local/lib/python3.9/dist-packages (from google-auth-
oauthlib<0.5,>=0.4.1->tensorboard->simpletransformers) (1.3.1)
Requirement already satisfied: zipp>=0.5 in
/usr/local/lib/python3.9/dist-packages (from importlib-metadata>=1.4-
>streamlit->simpletransformers) (3.15.0)
Collecting pygments<3.0.0,>=2.13.0
  Downloading Pygments-2.14.0-py3-none-any.whl (1.1 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 1.1/1.1 MB 80.6 MB/s eta
0:00:00
arkdown-it-py<3.0.0,>=2.2.0
  Downloading markdown_it_py-2.2.0-py3-none-any.whl (84 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 84.5/84.5 KB 14.8 MB/s eta
0:00:00
ent already satisfied: pytz-deprecation-shim in
/usr/local/lib/python3.9/dist-packages (from tzlocal>=1.1->streamlit-
>simpletransformers) (0.1.0.post0)
Requirement already satisfied: decorator>=3.4.0 in
/usr/local/lib/python3.9/dist-packages (from validators>=0.2-
>streamlit->simpletransformers) (4.4.2)
Requirement already satisfied: MarkupSafe>=2.1.1 in
/usr/local/lib/python3.9/dist-packages (from werkzeug>=1.0.1-
>tensorboard->simpletransformers) (2.1.2)
Collecting smmap<6,>=3.0.1
  Downloading smmap-5.0.0-py3-none-any.whl (24 kB)
Requirement already satisfied: pyrsistent!=0.17.0,!=0.17.1,!
=0.17.2,>=0.14.0 in /usr/local/lib/python3.9/dist-packages (from
jsonschema>=3.0->altair<5,>=3.2.0->streamlit->simpletransformers)
(0.19.3)
Collecting mdurl~=0.1
  Downloading mdurl-0.1.2-py3-none-any.whl (10.0 kB)
```

```
Requirement already satisfied: pyasn1<0.5.0,>=0.4.6 in
/usr/local/lib/python3.9/dist-packages (from pyasn1-modules>=0.2.1-
>google-auth<3,>=1.6.3->tensorboard->simpletransformers) (0.4.8)
Requirement already satisfied: oauthlib>=3.0.0 in
/usr/local/lib/python3.9/dist-packages (from requests-oauthlib>=0.7.0-
>google-auth-oauthlib<0.5,>=0.4.1->tensorboard->simpletransformers)
(3.2.2)
Requirement already satisfied: tzdata in
/usr/local/lib/python3.9/dist-packages (from pytz-deprecation-shim-
>tzlocal>=1.1->streamlit->simpletransformers) (2022.7)
Building wheels for collected packages: seqeval, validators, pathtools
  Building wheel for seqeval (setup.py) ... e=seqeval-1.2.2-py3-none-
any.whl size=16179
sha256=8832cd34ccc253a5bfc037bf1982c1d36ef71f82853054b145d9416b5d3d26f
7
  Stored in directory:
/root/.cache/pip/wheels/e2/a5/92/2c80d1928733611c2747a9820e1324a683552
4d9411510c142
  Building wheel for validators (setup.py) ... e=validators-0.20.0-
py3-none-any.whl size=19581
sha256=d717005201fdf37909950fda14a5f0602c3541dd5fdf62e35f730297d984e38
c
  Stored in directory:
/root/.cache/pip/wheels/2d/f0/a8/1094fca7a7e5d0d12ff56e0c64675d72aa5cc
81a5fc200e849
  Building wheel for pathtools (setup.py) ... e=pathtools-0.1.2-py3-
none-any.whl size=8807
sha256=cf8e0f1371ecc29e91db51566f52117b5087e0f63774a3f2b7e8e6e2083dec1
1
  Stored in directory:
/root/.cache/pip/wheels/b7/0a/67/ada2a22079218c75a88361c0782855cc72aeb
c4d18d0289d05
Successfully built seqeval validators pathtools
Installing collected packages: tokenizers, sentencepiece, pathtools,
appdirs, xxhash, watchdog, validators, smmap, setproctitle, sentry-
sdk, semver, pympler, pygments, multidict, mdurl, frozenlist, docker-
pycreds, dill, charset-normalizer, blinker, async-timeout, yarl,
responses, pydeck, multiprocess, markdown-it-py, huggingface-hub,
gitdb, aiosignal, transformers, seqeval, rich, GitPython, aiohttp,
wandb, streamlit, datasets, simpletransformers
  Attempting uninstall: pygments
    Found existing installation: Pygments 2.6.1
    Uninstalling Pygments-2.6.1:
      Successfully uninstalled Pygments-2.6.1
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.
ipython 7.9.0 requires jedi>=0.10, which is not installed.
Successfully installed GitPython-3.1.31 aiohttp-3.8.4 aiosignal-1.3.1
appdirs-1.4.4 async-timeout-4.0.2 blinker-1.5 charset-normalizer-3.1.0
```

```
datasets-2.10.1 dill-0.3.6 docker-pycreds-0.4.0 frozenlist-1.3.3
gitdb-4.0.10 huggingface-hub-0.13.2 markdown-it-py-2.2.0 mdurl-0.1.2
multidict-6.0.4 multiprocess-0.70.14 pathtools-0.1.2 pydeck-0.8.0
pygments-2.14.0 pympler-1.0.1 responses-0.18.0 rich-13.3.2 semver-
2.13.0 sentencepiece-0.1.97 sentry-sdk-1.16.0 seqeval-1.2.2
setproctitle-1.3.2 simpletransformers-0.63.9 smmap-5.0.0 streamlit-
1.20.0 tokenizers-0.13.2 transformers-4.26.1 validators-0.20.0 wandb-
0.13.11 watchdog-2.3.1 xxhash-3.2.0 yarl-1.8.2
```

```python
# Import libraries
import pandas as pd
import numpy as np
import sklearn
from sklearn.metrics import classification_report
from simpletransformers.classification import ClassificationModel,
ClassificationArgs
import matplotlib.pyplot as plt
import seaborn as sn

from sklearn.datasets import fetch_20newsgroups

# load only a sub-selection of the categories (4 in our case)
categories = ['alt.atheism', 'comp.graphics', 'sci.med', 'sci.space']

# remove the headers, footers and quotes (to avoid overfitting)
newsgroups_train = fetch_20newsgroups(subset='train',
remove=('headers', 'footers', 'quotes'), categories=categories,
random_state=42)
newsgroups_test = fetch_20newsgroups(subset='test', remove=('headers',
'footers', 'quotes'), categories=categories, random_state=42)

from collections import Counter
Counter(newsgroups_train.target)
```

```
Counter({3: 593, 1: 584, 2: 594, 0: 480})
```

```python
Counter(newsgroups_test.target)
```

```
Counter({1: 389, 2: 396, 0: 319, 3: 394})
```

```python
train = pd.DataFrame({'text': newsgroups_train.data, 'labels':
newsgroups_train.target})

print(len(train))
train.head(5)
```

```
2251

                                                text  labels
0  WHile we are on the subject of the shuttle sof...       3
1  There is a program called Graphic Workshop you...       1
2                                                          2
```

```
3  My girlfriend is in pain from kidney stones. S...       2
4  I think that's the correct spelling..\n\tI am ...       2
```

```
test = pd.DataFrame({'text': newsgroups_test.data, 'labels':
newsgroups_test.target})
```

```
print(len(test))
test.head(5)
```

```
1498
```

```
                                                text  labels
0  \nAnd guess who's here in your place.\n\nPleas...       1
1  Does anyone know if any of Currier and Ives et...       1
2  =FLAME ON\n=\n=Reading through the posts about...       2
3  \nBut in this case I said I hoped that BCCI wa...       0
4  \nIn the kind I have made I used a Lite sour c...       2
```

```
from sklearn.model_selection import train_test_split

train, dev = train_test_split(train, test_size=0.1, random_state=0,
                               stratify=train[['labels']])
```

```
print(len(train))
print("train:", train[['labels']].value_counts(sort=False))
train.head(3)
```

```
2025
train: labels
0         432
1         525
2         534
3         534
dtype: int64
```

```
                                                   text  labels
559   I wonder how many atheists out there care to s...       0
2060  We are interested in purchasing a grayscale pr...       1
1206  Dear Binary Newsers,\n\nI am looking for Quick...       1
```

```
print(len(dev))
print("dev:", dev[['labels']].value_counts(sort=False))
dev.head(3)
```

```
226
dev: labels
0         48
1         59
2         60
3         59
dtype: int64
```

```
                                                   text  labels
1570  I'd dump him.  Rude is rude and it seems he en...       2
1761  Hi Everyone ::\n\nI am  looking for  some soft...       1
455   A friend of mine has been diagnosed with Psori...       2
```

RoBERTa model

```
# Model configuration #
https://simpletransformers.ai/docs/usage/#configuring-a-simple-
transformers-model
model_args = ClassificationArgs()

model_args.overwrite_output_dir=True # overwrite existing saved models
in the same directory
model_args.evaluate_during_training=True # to perform evaluation while
training the model
# (eval data should be passed to the training method)

model_args.num_train_epochs=10 # number of epochs
model_args.train_batch_size=32 # batch size
model_args.learning_rate=4e-6 # learning rate
model_args.max_seq_length=256 # maximum sequence length
# Note! Increasing max_seq_len may provide better performance, but
training time will increase.
# For educational purposes, we set max_seq_len to 256.

# Early stopping to combat overfitting:
https://simpletransformers.ai/docs/tips-and-tricks/#using-early-
stopping
model_args.use_early_stopping=True
model_args.early_stopping_delta=0.01 # "The improvement over
best_eval_loss necessary to count as a better checkpoint"
model_args.early_stopping_metric='eval_loss'
model_args.early_stopping_metric_minimize=True
model_args.early_stopping_patience=2
model_args.evaluate_during_training_steps=32 # how

# Checking steps per epoch
steps_per_epoch = int(np.ceil(len(train) /
float(model_args.train_batch_size)))
print('Each epoch will have {:,} steps.'.format(steps_per_epoch)) # 64
steps = validating 2 times per epoch
```

```
Each epoch will have 64 steps.
```

```
model = ClassificationModel('roberta', 'roberta-base', num_labels=4,
args=model_args, use_cuda=True) # CUDA is enabled
```

```
loading configuration file config.json from cache at
/root/.cache/huggingface/hub/models--roberta-base/snapshots/bc2764f8af
2e92b6eb5679868df33e224075ca68/config.json
Model config RobertaConfig {
  "architectures": [
    "RobertaForMaskedLM"
  ],
  "attention_probs_dropout_prob": 0.1,
  "bos_token_id": 0,
  "classifier_dropout": null,
  "eos_token_id": 2,
  "hidden_act": "gelu",
  "hidden_dropout_prob": 0.1,
  "hidden_size": 768,
  "id2label": {
    "0": "LABEL_0",
    "1": "LABEL_1",
    "2": "LABEL_2",
    "3": "LABEL_3"
  },
  "initializer_range": 0.02,
  "intermediate_size": 3072,
  "label2id": {
    "LABEL_0": 0,
    "LABEL_1": 1,
    "LABEL_2": 2,
    "LABEL_3": 3
  },
  "layer_norm_eps": 1e-05,
  "max_position_embeddings": 514,
  "model_type": "roberta",
  "num_attention_heads": 12,
  "num_hidden_layers": 12,
  "pad_token_id": 1,
  "position_embedding_type": "absolute",
  "transformers_version": "4.26.1",
  "type_vocab_size": 1,
  "use_cache": true,
  "vocab_size": 50265
}

loading weights file pytorch_model.bin from cache at
/root/.cache/huggingface/hub/models--roberta-base/snapshots/bc2764f8af
2e92b6eb5679868df33e224075ca68/pytorch_model.bin
Some weights of the model checkpoint at roberta-base were not used
when initializing RobertaForSequenceClassification:
['lm_head.dense.weight', 'lm_head.decoder.weight',
'roberta.pooler.dense.bias', 'roberta.pooler.dense.weight',
'lm_head.dense.bias', 'lm_head.bias', 'lm_head.layer_norm.bias',
'lm_head.layer_norm.weight']
- This IS expected if you are initializing
RobertaForSequenceClassification from the checkpoint of a model
trained on another task or with another architecture (e.g.
initializing a BertForSequenceClassification model from a
```

```
BertForPreTraining model).
- This IS NOT expected if you are initializing
RobertaForSequenceClassification from the checkpoint of a model that
you expect to be exactly identical (initializing a
BertForSequenceClassification model from a
BertForSequenceClassification model).
Some weights of RobertaForSequenceClassification were not initialized
from the model checkpoint at roberta-base and are newly initialized:
['classifier.out_proj.bias', 'classifier.dense.weight',
'classifier.out_proj.weight', 'classifier.dense.bias']
You should probably TRAIN this model on a down-stream task to be able
to use it for predictions and inference.
loading file vocab.json from cache at
/root/.cache/huggingface/hub/models--roberta-base/snapshots/bc2764f8af
2e92b6eb5679868df33e224075ca68/vocab.json
loading file merges.txt from cache at
/root/.cache/huggingface/hub/models--roberta-base/snapshots/bc2764f8af
2e92b6eb5679868df33e224075ca68/merges.txt
loading file tokenizer.json from cache at
/root/.cache/huggingface/hub/models--roberta-base/snapshots/bc2764f8af
2e92b6eb5679868df33e224075ca68/tokenizer.json
loading file added_tokens.json from cache at None
loading file special_tokens_map.json from cache at None
loading file tokenizer_config.json from cache at None
loading configuration file config.json from cache at
/root/.cache/huggingface/hub/models--roberta-base/snapshots/bc2764f8af
2e92b6eb5679868df33e224075ca68/config.json
Model config RobertaConfig {
  "_name_or_path": "roberta-base",
  "architectures": [
    "RobertaForMaskedLM"
  ],
  "attention_probs_dropout_prob": 0.1,
  "bos_token_id": 0,
  "classifier_dropout": null,
  "eos_token_id": 2,
  "hidden_act": "gelu",
  "hidden_dropout_prob": 0.1,
  "hidden_size": 768,
  "initializer_range": 0.02,
  "intermediate_size": 3072,
  "layer_norm_eps": 1e-05,
  "max_position_embeddings": 514,
  "model_type": "roberta",
  "num_attention_heads": 12,
  "num_hidden_layers": 12,
  "pad_token_id": 1,
  "position_embedding_type": "absolute",
  "transformers_version": "4.26.1",
  "type_vocab_size": 1,
  "use_cache": true,
  "vocab_size": 50265
}
```

```
print(str(model.args).replace(',', '\n')) # model args
```

```
ClassificationArgs(adafactor_beta1=None
 adafactor_clip_threshold=1.0
 adafactor_decay_rate=-0.8
 adafactor_eps=(1e-30
 0.001)
 adafactor_relative_step=True
 adafactor_scale_parameter=True
 adafactor_warmup_init=True
 adam_betas=(0.9
 0.999)
 adam_epsilon=1e-08
 best_model_dir='outputs/best_model'
 cache_dir='cache_dir/'
 config={}
 cosine_schedule_num_cycles=0.5
 custom_layer_parameters=[]
 custom_parameter_groups=[]
 dataloader_num_workers=0
 do_lower_case=False
 dynamic_quantize=False
 early_stopping_consider_epochs=False
 early_stopping_delta=0.01
 early_stopping_metric='eval_loss'
 early_stopping_metric_minimize=True
 early_stopping_patience=2
 encoding=None
 eval_batch_size=8
 evaluate_during_training=True
 evaluate_during_training_silent=True
 evaluate_during_training_steps=32
 evaluate_during_training_verbose=False
 evaluate_each_epoch=True
 fp16=True
 gradient_accumulation_steps=1
 learning_rate=4e-06
 local_rank=-1
 logging_steps=50
 loss_type=None
 loss_args={}
 manual_seed=None
 max_grad_norm=1.0
 max_seq_length=256
 model_name='roberta-base'
```

```
model_type='roberta'
multiprocessing_chunksize=-1
n_gpu=1
no_cache=False
no_save=False
not_saved_args=[]
num_train_epochs=10
optimizer='AdamW'
output_dir='outputs/'
overwrite_output_dir=True
polynomial_decay_schedule_lr_end=1e-07
polynomial_decay_schedule_power=1.0
process_count=1
quantized_model=False
reprocess_input_data=True
save_best_model=True
save_eval_checkpoints=True
save_model_every_epoch=True
save_optimizer_and_scheduler=True
save_steps=2000
scheduler='linear_schedule_with_warmup'
silent=False
skip_special_tokens=True
tensorboard_dir=None
thread_count=None
tokenizer_name='roberta-base'
tokenizer_type=None
train_batch_size=32
train_custom_parameters_only=False
use_cached_eval_features=False
use_early_stopping=True
use_hf_datasets=False
use_multiprocessing=True
use_multiprocessing_for_evaluation=True
wandb_kwargs={}
wandb_project=None
warmup_ratio=0.06
warmup_steps=0
weight_decay=0.0
model_class='ClassificationModel'
labels_list=[0
1
2
3]
labels_map={}
lazy_delimiter='\t'
lazy_labels_column=1
lazy_loading=False
lazy_loading_start_line=1
lazy_text_a_column=None
```

```
 lazy_text_b_column=None
 lazy_text_column=0
 onnx=False
 regression=False
 sliding_window=False
 special_tokens_list=[]
 stride=0.8
 tie_value=1)
```

```
_, history = model.train_model(train, eval_df=dev)
```

```
{"model_id":"7ebcad96aab14db591a63ebd71b7bc01","version_major":2,"vers
ion_minor":0}
```

```
{"model_id":"a7c824260e4543edb75d35f24f0a1793","version_major":2,"vers
ion_minor":0}
```

```
{"model_id":"74a6ad80579f4c0289b4747114c7502e","version_major":2,"vers
ion_minor":0}
```

```
{"model_id":"733b118c4efc4f19ae5f9eda73b3f837","version_major":2,"vers
ion_minor":0}
```

```
Configuration saved in outputs/checkpoint-32/config.json
Model weights saved in outputs/checkpoint-32/pytorch_model.bin
tokenizer config file saved in
outputs/checkpoint-32/tokenizer_config.json
Special tokens file saved in
outputs/checkpoint-32/special_tokens_map.json
Configuration saved in outputs/best_model/config.json
Model weights saved in outputs/best_model/pytorch_model.bin
tokenizer config file saved in
outputs/best_model/tokenizer_config.json
Special tokens file saved in
outputs/best_model/special_tokens_map.json
```

```
{"model_id":"6464b368f2a545b6a2c52bd6709dd8c9","version_major":2,"vers
ion_minor":0}
```

```
Configuration saved in outputs/checkpoint-64/config.json
Model weights saved in outputs/checkpoint-64/pytorch_model.bin
tokenizer config file saved in
outputs/checkpoint-64/tokenizer_config.json
Special tokens file saved in
outputs/checkpoint-64/special_tokens_map.json
Configuration saved in outputs/best_model/config.json
Model weights saved in outputs/best_model/pytorch_model.bin
tokenizer config file saved in
outputs/best_model/tokenizer_config.json
Special tokens file saved in
outputs/best_model/special_tokens_map.json
Configuration saved in outputs/checkpoint-64-epoch-1/config.json
```

```
Model weights saved in outputs/checkpoint-64-epoch-1/pytorch_model.bin
tokenizer config file saved in
outputs/checkpoint-64-epoch-1/tokenizer_config.json
Special tokens file saved in
outputs/checkpoint-64-epoch-1/special_tokens_map.json
```

```
{"model_id":"3fd9b58a33524e48a61c10c25c8c9d2f","version_major":2,"version_minor":0}
```

```
{"model_id":"aa98ab728e3343799b9c939c12e3571e","version_major":2,"version_minor":0}
```

```
{"model_id":"242b868de3b241e98968b9355a3d2a7c","version_major":2,"version_minor":0}
```

```
Configuration saved in outputs/checkpoint-96/config.json
Model weights saved in outputs/checkpoint-96/pytorch_model.bin
tokenizer config file saved in
outputs/checkpoint-96/tokenizer_config.json
Special tokens file saved in
outputs/checkpoint-96/special_tokens_map.json
Configuration saved in outputs/best_model/config.json
Model weights saved in outputs/best_model/pytorch_model.bin
tokenizer config file saved in
outputs/best_model/tokenizer_config.json
Special tokens file saved in
outputs/best_model/special_tokens_map.json
```

```
{"model_id":"076442fcc09e4f48abb8b670122b1fa1","version_major":2,"version_minor":0}
```

```
Configuration saved in outputs/checkpoint-128/config.json
Model weights saved in outputs/checkpoint-128/pytorch_model.bin
tokenizer config file saved in
outputs/checkpoint-128/tokenizer_config.json
Special tokens file saved in
outputs/checkpoint-128/special_tokens_map.json
Configuration saved in outputs/best_model/config.json
Model weights saved in outputs/best_model/pytorch_model.bin
tokenizer config file saved in
outputs/best_model/tokenizer_config.json
Special tokens file saved in
outputs/best_model/special_tokens_map.json
Configuration saved in outputs/checkpoint-128-epoch-2/config.json
Model weights saved in
outputs/checkpoint-128-epoch-2/pytorch_model.bin
tokenizer config file saved in
outputs/checkpoint-128-epoch-2/tokenizer_config.json
Special tokens file saved in
outputs/checkpoint-128-epoch-2/special_tokens_map.json
```

```
{"model_id":"f0eaf99ed59b4428ab31188543ab0271","version_major":2,"version_minor":0}
```

```
{"model_id":"b1362237eb094ea6a7079f90e7fec096","version_major":2,"version_minor":0}
```

```
{"model_id":"3efff51614394bc7ab46f2b47013e047","version_major":2,"version_minor":0}
```

```
Configuration saved in outputs/checkpoint-160/config.json
Model weights saved in outputs/checkpoint-160/pytorch_model.bin
tokenizer config file saved in
outputs/checkpoint-160/tokenizer_config.json
Special tokens file saved in
outputs/checkpoint-160/special_tokens_map.json
Configuration saved in outputs/best_model/config.json
Model weights saved in outputs/best_model/pytorch_model.bin
tokenizer config file saved in
outputs/best_model/tokenizer_config.json
Special tokens file saved in
outputs/best_model/special_tokens_map.json
```

```
{"model_id":"ca9741711ea04327bac4ecad59ddcb98","version_major":2,"version_minor":0}
```

```
Configuration saved in outputs/checkpoint-192/config.json
Model weights saved in outputs/checkpoint-192/pytorch_model.bin
tokenizer config file saved in
outputs/checkpoint-192/tokenizer_config.json
Special tokens file saved in
outputs/checkpoint-192/special_tokens_map.json
Configuration saved in outputs/best_model/config.json
Model weights saved in outputs/best_model/pytorch_model.bin
tokenizer config file saved in
outputs/best_model/tokenizer_config.json
Special tokens file saved in
outputs/best_model/special_tokens_map.json
Configuration saved in outputs/checkpoint-192-epoch-3/config.json
Model weights saved in
outputs/checkpoint-192-epoch-3/pytorch_model.bin
tokenizer config file saved in
outputs/checkpoint-192-epoch-3/tokenizer_config.json
Special tokens file saved in
outputs/checkpoint-192-epoch-3/special_tokens_map.json
```

```
{"model_id":"04a144a650da4b019b3b2ccd3c12ff5a","version_major":2,"version_minor":0}
```

```
{"model_id":"e3a17a009d814c6ba87a0d226f68d047","version_major":2,"version_minor":0}
```

```
{"model_id":"8abbb018b9f94a5ab2ccc264df2d2fa0","version_major":2,"vers
ion_minor":0}
```

```
Configuration saved in outputs/checkpoint-224/config.json
Model weights saved in outputs/checkpoint-224/pytorch_model.bin
tokenizer config file saved in
outputs/checkpoint-224/tokenizer_config.json
Special tokens file saved in
outputs/checkpoint-224/special_tokens_map.json
```

```
{"model_id":"1c1a9c1404e740c38f4810f8d1532f9e","version_major":2,"vers
ion_minor":0}
```

```
Configuration saved in outputs/checkpoint-256/config.json
Model weights saved in outputs/checkpoint-256/pytorch_model.bin
tokenizer config file saved in
outputs/checkpoint-256/tokenizer_config.json
Special tokens file saved in
outputs/checkpoint-256/special_tokens_map.json
Configuration saved in outputs/best_model/config.json
Model weights saved in outputs/best_model/pytorch_model.bin
tokenizer config file saved in
outputs/best_model/tokenizer_config.json
Special tokens file saved in
outputs/best_model/special_tokens_map.json
Configuration saved in outputs/checkpoint-256-epoch-4/config.json
Model weights saved in
outputs/checkpoint-256-epoch-4/pytorch_model.bin
tokenizer config file saved in
outputs/checkpoint-256-epoch-4/tokenizer_config.json
Special tokens file saved in
outputs/checkpoint-256-epoch-4/special_tokens_map.json
```

```
{"model_id":"2fde136046b9432bad727ff88de83ed4","version_major":2,"vers
ion_minor":0}
```

```
{"model_id":"16b83a1c81fe4bc9b6d2a186f1346f4b","version_major":2,"vers
ion_minor":0}
```

```
{"model_id":"ed98775e244b4cfaa8b0c4f57d9ff7e9","version_major":2,"vers
ion_minor":0}
```

```
Configuration saved in outputs/checkpoint-288/config.json
Model weights saved in outputs/checkpoint-288/pytorch_model.bin
tokenizer config file saved in
outputs/checkpoint-288/tokenizer_config.json
Special tokens file saved in
outputs/checkpoint-288/special_tokens_map.json
Configuration saved in outputs/best_model/config.json
Model weights saved in outputs/best_model/pytorch_model.bin
tokenizer config file saved in
outputs/best_model/tokenizer_config.json
```

```
Special tokens file saved in
outputs/best_model/special_tokens_map.json
```

{"model_id":"258b725e475345e28503a05a2e291e42","version_major":2,"version_minor":0}

```
Configuration saved in outputs/checkpoint-320/config.json
Model weights saved in outputs/checkpoint-320/pytorch_model.bin
tokenizer config file saved in
outputs/checkpoint-320/tokenizer_config.json
Special tokens file saved in
outputs/checkpoint-320/special_tokens_map.json
Configuration saved in outputs/checkpoint-320-epoch-5/config.json
Model weights saved in
outputs/checkpoint-320-epoch-5/pytorch_model.bin
tokenizer config file saved in
outputs/checkpoint-320-epoch-5/tokenizer_config.json
Special tokens file saved in
outputs/checkpoint-320-epoch-5/special_tokens_map.json
```

{"model_id":"cf0f7a0deb984436afa0a95c22b4cdf2","version_major":2,"version_minor":0}

{"model_id":"6a5e23e929d9494591945340a6ce1fd3","version_major":2,"version_minor":0}

{"model_id":"cb021e24dc124eb3837b8c9eaad0240d","version_major":2,"version_minor":0}

```
Configuration saved in outputs/checkpoint-352/config.json
Model weights saved in outputs/checkpoint-352/pytorch_model.bin
tokenizer config file saved in
outputs/checkpoint-352/tokenizer_config.json
Special tokens file saved in
outputs/checkpoint-352/special_tokens_map.json
```

{"model_id":"679a6772ebaa4217bd3305be0b2796a3","version_major":2,"version_minor":0}

```
Configuration saved in outputs/checkpoint-384/config.json
Model weights saved in outputs/checkpoint-384/pytorch_model.bin
tokenizer config file saved in
outputs/checkpoint-384/tokenizer_config.json
Special tokens file saved in
outputs/checkpoint-384/special_tokens_map.json
Configuration saved in outputs/best_model/config.json
Model weights saved in outputs/best_model/pytorch_model.bin
tokenizer config file saved in
outputs/best_model/tokenizer_config.json
Special tokens file saved in
outputs/best_model/special_tokens_map.json
Configuration saved in outputs/checkpoint-384-epoch-6/config.json
```

```
Model weights saved in
outputs/checkpoint-384-epoch-6/pytorch_model.bin
tokenizer config file saved in
outputs/checkpoint-384-epoch-6/tokenizer_config.json
Special tokens file saved in
outputs/checkpoint-384-epoch-6/special_tokens_map.json
```

```
{"model_id":"a2067e13fa37401099e6a052a428d45d","version_major":2,"vers
ion_minor":0}
```

```
{"model_id":"f30a3af77b4f41b0812c826442202de3","version_major":2,"vers
ion_minor":0}
```

```
{"model_id":"b5c616ac2bed4c05add8c61fedecf245","version_major":2,"vers
ion_minor":0}
```

```
Configuration saved in outputs/checkpoint-416/config.json
Model weights saved in outputs/checkpoint-416/pytorch_model.bin
tokenizer config file saved in
outputs/checkpoint-416/tokenizer_config.json
Special tokens file saved in
outputs/checkpoint-416/special_tokens_map.json
```

```
{"model_id":"c6690ea4129741368106957b361a52ec","version_major":2,"vers
ion_minor":0}
```

```
Configuration saved in outputs/checkpoint-448/config.json
Model weights saved in outputs/checkpoint-448/pytorch_model.bin
tokenizer config file saved in
outputs/checkpoint-448/tokenizer_config.json
Special tokens file saved in
outputs/checkpoint-448/special_tokens_map.json
Configuration saved in outputs/checkpoint-448-epoch-7/config.json
Model weights saved in
outputs/checkpoint-448-epoch-7/pytorch_model.bin
tokenizer config file saved in
outputs/checkpoint-448-epoch-7/tokenizer_config.json
Special tokens file saved in
outputs/checkpoint-448-epoch-7/special_tokens_map.json
```

```
{"model_id":"3426a6285cbf4cef8d37777e1d51c86d","version_major":2,"vers
ion_minor":0}
```

```
{"model_id":"1034d195a84a47e189d98a819b1d4bce","version_major":2,"vers
ion_minor":0}
```

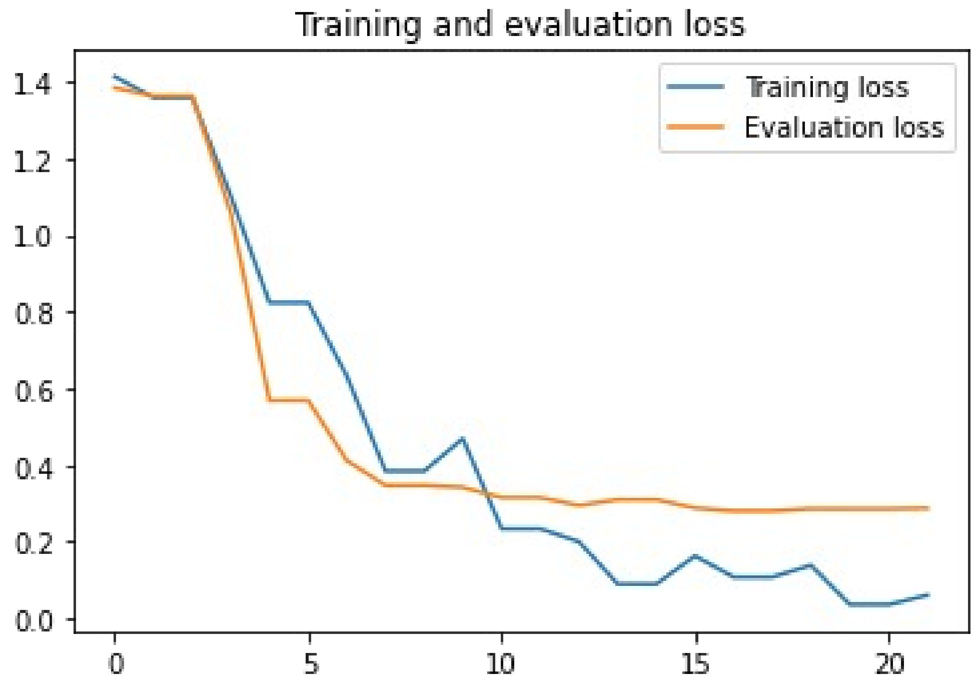
```
{"model_id":"42f27f92fa284a25ac7b105acc6574cc","version_major":2,"vers
ion_minor":0}
```

```
Configuration saved in outputs/checkpoint-480/config.json
Model weights saved in outputs/checkpoint-480/pytorch_model.bin
tokenizer config file saved in
```

```
outputs/checkpoint-480/tokenizer_config.json
Special tokens file saved in
outputs/checkpoint-480/special_tokens_map.json
Configuration saved in outputs/config.json
Model weights saved in outputs/pytorch_model.bin
tokenizer config file saved in outputs/tokenizer_config.json
Special tokens file saved in outputs/special_tokens_map.json
```

```
# Training and evaluation loss
train_loss = history['train_loss']
eval_loss = history['eval_loss']
plt.plot(train_loss, label='Training loss')
plt.plot(eval_loss, label='Evaluation loss')
plt.title('Training and evaluation loss')
plt.legend()
```

```
<matplotlib.legend.Legend at 0x7fbde018ae50>
```



```
# Evaluate the model
result, model_outputs, wrong_predictions = model.eval_model(dev)
result
```

{"model_id":"2e91d34ce6dd4fcbbc1d84905f69d8aa","version_major":2,"version_minor":0}

{"model_id":"52947dc27d41408ebe2dd040adcc3fa6","version_major":2,"version_minor":0}

```
{'mcc': 0.8407842504687129, 'eval_loss': 0.2880030007197939}
```

```
predicted, probabilities = model.predict(test.text.to_list())
test['predicted'] = predicted
```

```
{"model_id":"d06117f0359145ed9f4272f1a319fa0e","version_major":2,"vers
ion_minor":0}
```

```
{"model_id":"fadc6ab7ef064e91989d9c7a731bc97d","version_major":2,"vers
ion_minor":0}
```

```
test.head(5)
```

```
                                                text  labels
predicted
0  \nAnd guess who's here in your place.\n\nPleas...       1
1
1  Does anyone know if any of Currier and Ives et...       1
1
2  =FLAME ON\n=\n=Reading through the posts about...       2
0
3  \nBut in this case I said I hoped that BCCI wa...       0
0
4  \nIn the kind I have made I used a Lite sour c...       2
2
```

```
# Result (note: your result can be different due to randomness in
operations)
print(classification_report(test['labels'], test['predicted']))
```

```
              precision    recall  f1-score   support

           0       0.82      0.81      0.82       319
           1       0.90      0.91      0.90       389
           2       0.89      0.88      0.89       396
           3       0.82      0.83      0.82       394

    accuracy                           0.86      1498
   macro avg       0.86      0.86      0.86      1498
weighted avg       0.86      0.86      0.86      1498
```

Both models seem to perform well in terms of accuracy, with an accuracy score of 0.86 and 0.87 respectively. However, when comparing the results of the two models, it seems that the second model performs slightly better than the first one.

This is mainly due to the fact that the second model has slightly higher precision and recall scores across all classes, as well as a higher F1-score. In particular, the second model has higher precision scores for classes 0 and 2, higher recall scores for classes 1 and 3, and higher F1-scores for classes 1 and 2.

Therefore, if we consider precision and recall as important metrics, we would conclude that the second model performs slightly better than the first one.

When we compare roBERTa model and SVM using bag-of-words model regarding their performance using precision, recall metrics, roBERTa tends to have higher precision score than SVM using bag-of-words thanks to its ability to learn more complex links between words and their contexts. Whereas SVM using bag-of-words may have higher recall score than the other model because it is more focused on individual words and their frequency in the text considering recall score shows the percentage of examples that are successfully categorized out of all the instances that fall under a given class.