

```
%cd '/content/drive/MyDrive/depression-detection-lt-edi-2022/models'

/content/drive/.shortcut-targets-by-id/1xD1ZlQhDatgu8BZrlnav5pFJeOkI8nOi/depression-detection-lt-edi-2022/model

from google.colab import drive
drive.mount('/content/drive')

Mounted at /content/drive

import logging
logging.getLogger("transformers").setLevel(logging.ERROR)

import torch
torch.cuda.empty_cache()

# Confirm that the GPU is detected

assert torch.cuda.is_available()

# Get the GPU device name.
device_name = torch.cuda.get_device_name()
n_gpu = torch.cuda.device_count()
print(f"Found device: {device_name}, n_gpu: {n_gpu}")
device = torch.device("cuda")

Found device: Tesla T4, n_gpu: 1

!pip install transformers simpletransformers pandas==1.2.5 scikit-learn==0.23.1 tqdm==4.62.3
!pip install transformers
!pip install -U -q PyDrive

Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: transformers in /usr/local/lib/python3.10/dist-packages (4.28.1)
Collecting simpletransformers
  Downloading simpletransformers-0.63.11-py3-none-any.whl (250 kB)
    250.7/250.7 kB 5.7 MB/s eta 0:00:00
Collecting pandas==1.2.5
  Using cached pandas-1.2.5.tar.gz (5.5 MB)
  Installing build dependencies ... done
  Getting requirements to build wheel ... done
  Preparing metadata (pyproject.toml) ... done
Collecting scikit-learn==0.23.1
  Using cached scikit-learn-0.23.1.tar.gz (7.2 MB)
error: subprocess-exited-with-error

  × pip subprocess to install build dependencies did not run successfully.
  | exit code: 1
  |_ See above for output.

note: This error originates from a subprocess, and is likely not a problem with pip.
Installing build dependencies ... error
error: subprocess-exited-with-error

  × pip subprocess to install build dependencies did not run successfully.
  | exit code: 1
  |_ See above for output.

note: This error originates from a subprocess, and is likely not a problem with pip.
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: transformers in /usr/local/lib/python3.10/dist-packages (4.28.1)
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-packages (from transformers) (3.12.0)
Requirement already satisfied: huggingface-hub<1.0,>=0.11.0 in /usr/local/lib/python3.10/dist-packages (from transformers) (0.12.1)
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.10/dist-packages (from transformers) (1.24.3)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from transformers) (23.1)
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.10/dist-packages (from transformers) (6.0.1)
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.10/dist-packages (from transformers) (2023.10.3)
Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-packages (from transformers) (2.27.1)
Requirement already satisfied: tokenizers!=0.11.3,<0.14,>=0.11.1 in /usr/local/lib/python3.10/dist-packages (from transformers) (0.13.3)
Requirement already satisfied: tqdm>=4.27 in /usr/local/lib/python3.10/dist-packages (from transformers) (4.65.0)
Requirement already satisfied: fsspec in /usr/local/lib/python3.10/dist-packages (from huggingface-hub<1.0,>=0.11.0) (2023.10.3)
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.10/dist-packages (from huggingface-hub<1.0,>=0.11.0) (4.5.0)
```

Requirement already satisfied: urllib3<1.27,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests-
 Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests->tr
 Requirement already satisfied: charset-normalizer~=2.0.0 in /usr/local/lib/python3.10/dist-packages (from requ
 Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests->transfor

```
!python preprocess_dataset.py
```

```
python3: can't open file '/content/drive/.shortcut-targets-by-id/1xD1ZIqHdatgu8BZr1nav5pFJeOkI8nOi/depression-c
```

```
!python reddit_depression_corpora.py
```

```
100% 50/50 [01:37<00:00, 1.96s/it]
depression: 72916 (27.1%)
suicidewatch: 34747 (12.9%)
lonely: 20105 (7.5%)
socialanxiety: 19697 (7.3%)
fitness: 10000 (3.7%)
EDAnonymous: 10000 (3.7%)
guns: 10000 (3.7%)
jokes: 10000 (3.7%)
legaladvice: 10000 (3.7%)
mentalhealth: 10000 (3.7%)
personalfinance: 10000 (3.7%)
parenting: 10000 (3.7%)
relationships: 10000 (3.7%)
meditation: 7852 (2.9%)
schizophrenia: 6660 (2.5%)
healthanxiety: 5436 (2.0%)
teaching: 3516 (1.3%)
alcoholism: 3087 (1.1%)
divorce: 2745 (1.0%)
ptsd: 2097 (0.8%)

Unique texts: 268858 (train: 263480 / validation: 5378).
```

```
%cd ..
```

```
/content/drive/.shortcut-targets-by-id/1xD1ZIqHdatgu8BZr1nav5pFJeOkI8nOi/depression-detection-lt-edi-2022
```

```
%cd models
```

```
/content/drive/.shortcut-targets-by-id/1xD1ZIqHdatgu8BZr1nav5pFJeOkI8nOi/depression-detection-lt-edi-2022/model
```

```
!python models_list.py
```

```
!pip install simpletransformers
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Collecting simpletransformers
  Using cached simpletransformers-0.63.11-py3-none-any.whl (250 kB)
Requirement already satisfied: numpy in /usr/local/lib/python3.10/dist-packages (from simpletransformers) (1.22
Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-packages (from simpletransformers) (2
Requirement already satisfied: tqdm>=4.47.0 in /usr/local/lib/python3.10/dist-packages (from simpletransformers
Requirement already satisfied: regex in /usr/local/lib/python3.10/dist-packages (from simpletransformers) (2022
Requirement already satisfied: transformers>=4.6.0 in /usr/local/lib/python3.10/dist-packages (from simpletrans
Collecting datasets (from simpletransformers)
  Downloading datasets-2.12.0-py3-none-any.whl (474 kB)
    ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 474.6/474.6 kB 9.6 MB/s eta 0:00:00
Requirement already satisfied: scipy in /usr/local/lib/python3.10/dist-packages (from simpletransformers) (1.10
Requirement already satisfied: scikit-learn in /usr/local/lib/python3.10/dist-packages (from simpletransformers
Collecting segeval (from simpletransformers)
  Downloading segeval-1.2.2.tar.gz (43 kB)
    ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 43.6/43.6 kB 5.0 MB/s eta 0:00:00
  Preparing metadata (setup.py) ... done
Requirement already satisfied: tensorboard in /usr/local/lib/python3.10/dist-packages (from simpletransformers)
Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages (from simpletransformers) (1.5
Requirement already satisfied: tokenizers in /usr/local/lib/python3.10/dist-packages (from simpletransformers)
Collecting wandb>=0.10.32 (from simpletransformers)
  Downloading wandb-0.15.0-py3-none-any.whl (2.0 MB)
    ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 2.0/2.0 MB 44.0 MB/s eta 0:00:00
Collecting streamlit (from simpletransformers)
  Downloading streamlit-1.22.0-py2.py3-none-any.whl (8.9 MB)
```

```

8.9/8.9 MB 66.0 MB/s eta 0:00:00
Collecting sentencepiece (from simpletransformers)
  Downloading sentencepiece-0.1.98-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (1.3 MB)
    1.3/1.3 MB 63.8 MB/s eta 0:00:00
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-packages (from transformers>=4.6.0->simpletransformers)
Requirement already satisfied: huggingface-hub<1.0,>=0.11.0 in /usr/local/lib/python3.10/dist-packages (from transformers>=4.6.0->simpletransformers)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from transformers>=4.6.0->simpletransformers)
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.10/dist-packages (from transformers>=4.6.0->simpletransformers)
Requirement already satisfied: Click!=8.0.0,>=7.0 in /usr/local/lib/python3.10/dist-packages (from wandb>=0.10.32->simpletransformers)
Collecting GitPython!=3.1.29,>=1.0.0 (from wandb>=0.10.32->simpletransformers)
  Downloading GitPython-3.1.31-py3-none-any.whl (184 kB)
    184.3/184.3 kB 22.3 MB/s eta 0:00:00
Requirement already satisfied: psutil>=5.0.0 in /usr/local/lib/python3.10/dist-packages (from wandb>=0.10.32->simpletransformers)
Collecting sentry-sdk>=1.0.0 (from wandb>=0.10.32->simpletransformers)
  Downloading sentry_sdk-1.21.1-py2.py3-none-any.whl (201 kB)
    201.7/201.7 kB 25.1 MB/s eta 0:00:00
Collecting docker-pycreds>=0.4.0 (from wandb>=0.10.32->simpletransformers)
  Downloading docker_pycreds-0.4.0-py2.py3-none-any.whl (9.0 kB)
Collecting pathtools (from wandb>=0.10.32->simpletransformers)
  Downloading pathtools-0.1.2.tar.gz (11 kB)
  Preparing metadata (setup.py) ... done
Collecting setproctitle (from wandb>=0.10.32->simpletransformers)
  Downloading setproctitle-1.3.2-cp310-cp310-manylinux_2_5_x86_64.manylinux1_x86_64.manylinux2014_x86_64.manylinux2017_x86_64.manylinux2018_x86_64.whl (28 kB)
Requirement already satisfied: setuptools in /usr/local/lib/python3.10/dist-packages (from wandb>=0.10.32->simpletransformers)
Requirement already satisfied: appdirs>=1.4.3 in /usr/local/lib/python3.10/dist-packages (from wandb>=0.10.32->simpletransformers)
Requirement already satisfied: protobuf!=4.21.0, <5, >=3.19.0 in /usr/local/lib/python3.10/dist-packages (from wandb>=0.10.32->simpletransformers)
Requirement already satisfied: urllib3<1.27, >=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests>=2.28.1->simpletransformers)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests>=2.28.1->simpletransformers)
Requirement already satisfied: charset-normalizer~2.0.0 in /usr/local/lib/python3.10/dist-packages (from requests>=2.28.1->simpletransformers)
Requirement already satisfied: idna<4, >=2.5 in /usr/local/lib/python3.10/dist-packages (from requests>=2.28.1->simpletransformers)
Requirement already satisfied: pyarrow>=8.0.0 in /usr/local/lib/python3.10/dist-packages (from datasets>=2.12.0->simpletransformers)
Collecting dill<0.3.7, >=0.3.0 (from datasets>=2.12.0->simpletransformers)
  Downloading dill-0.3.6-py3-none-any.whl (110 kB)

!python transformers_models.py

!python finetune.py

```

```
Epochs 1/2. Running Loss: 0.7875: 99% 247/250 [02:03<00:01, 2.89it/s]
Epochs 1/2. Running Loss: 0.7925: 99% 247/250 [02:03<00:01, 2.89it/s]
Epochs 1/2. Running Loss: 0.7925: 99% 248/250 [02:03<00:00, 2.89it/s]
Epochs 1/2. Running Loss: 0.7525: 99% 248/250 [02:03<00:00, 2.89it/s]
Epochs 1/2. Running Loss: 0.7525: 100% 249/250 [02:03<00:00, 2.89it/s]
Epochs 1/2. Running Loss: 0.9430: 100% 249/250 [02:04<00:00, 2.89it/s]
```

```
0% 0/600 [00:00<?, ?it/s]
```

```
0% 1/600 [00:01<11:10, 1.12s/it]
```

```
0% 2/600 [00:01<06:41, 1.49it/s]
```

```
Epochs 1/2. Running Loss: 0.9430: 100% 250/250 [02:19<00:00, 1.80it/s]
```

```
0% 0/600 [00:00<?, ?it/s]
```

```
0% 1/600 [00:01<17:58, 1.80s/it]
```

```
0% 2/600 [00:01<09:29, 1.05it/s]
```

```
Epoch 2 of 2: 100% 2/2 [05:39<00:00, 169.80s/it]
```

```
!python eval.py
```

```
2023-05-02 03:07:16.954594: I tensorflow/core/platform/cpu_feature_guard.cc:182] This TensorFlow binary is optimized
To enable the following instructions: AVX2 FMA, in other operations, rebuild TensorFlow with the appropriate compiler flags
2023-05-02 03:07:19.625166: W tensorflow/compiler/tf2tensorrt/utils/py_utils.cc:38] TF-TRT Warning: Could not find TensorRT
Evaluating roberta roberta-large v1
0% 2/600 [00:02<11:17, 1.13s/it]
100% 12/12 [00:09<00:00, 1.30it/s]
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1344: UndefinedMetricWarning: Precision is
_warn_prf(average, modifier, msg_start, len(result))
/content/drive/.shortcut-targets-by-id/1xD1ZlQhDatgu8BZrlnav5pFJeOkI8nOi/depression-detection-lt-edi-2022/model
self.results = self.results.append(result, ignore_index=True)
Evaluating roberta roberta-large v2
0% 2/600 [00:02<10:56, 1.10s/it]
100% 12/12 [00:09<00:00, 1.31it/s]
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1344: UndefinedMetricWarning: Precision is
_warn_prf(average, modifier, msg_start, len(result))
/content/drive/.shortcut-targets-by-id/1xD1ZlQhDatgu8BZrlnav5pFJeOkI8nOi/depression-detection-lt-edi-2022/model
self.results = self.results.append(result, ignore_index=True)
roberta-large & 0.563 & 0.399 & 0.387 & 0.353 \\
Evaluating bert bert-base-multilingual-cased v1
0% 2/600 [00:01<05:57, 1.67it/s]
100% 12/12 [00:03<00:00, 3.85it/s]
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1344: UndefinedMetricWarning: Precision is
_warn_prf(average, modifier, msg_start, len(result))
/content/drive/.shortcut-targets-by-id/1xD1ZlQhDatgu8BZrlnav5pFJeOkI8nOi/depression-detection-lt-edi-2022/model
self.results = self.results.append(result, ignore_index=True)
Evaluating bert bert-base-multilingual-cased v2
0% 2/600 [00:01<06:00, 1.66it/s]
100% 12/12 [00:03<00:00, 3.82it/s]
/content/drive/.shortcut-targets-by-id/1xD1ZlQhDatgu8BZrlnav5pFJeOkI8nOi/depression-detection-lt-edi-2022/model
self.results = self.results.append(result, ignore_index=True)
bert-base-multilingual-cased & 0.568 & 0.529 & 0.519 & 0.504 \\
Traceback (most recent call last):
  File "/content/drive/.shortcut-targets-by-id/1xD1ZlQhDatgu8BZrlnav5pFJeOkI8nOi/depression-detection-lt-edi-2022/eval.py", line 100, in <module>
    print(agg.get_result())
  File "/content/drive/.shortcut-targets-by-id/1xD1ZlQhDatgu8BZrlnav5pFJeOkI8nOi/depression-detection-lt-edi-2022/eval.py", line 100, in <module>
    values = self._get_best_result_by_main_metric(main_metric)
  File "/content/drive/.shortcut-targets-by-id/1xD1ZlQhDatgu8BZrlnav5pFJeOkI8nOi/depression-detection-lt-edi-2022/eval.py", line 100, in <module>
    return self.results.iloc[self.results[main_metric].idxmax()]
  File "/usr/local/lib/python3.10/dist-packages/pandas/core/series.py", line 2564, in idxmax
    i = self.argmax(axis, skipna, *args, **kwargs)
  File "/usr/local/lib/python3.10/dist-packages/pandas/core/base.py", line 655, in argmax
    return nanops.nanargmax(self, # type: ignore[return-value])
  File "/usr/local/lib/python3.10/dist-packages/pandas/core/nanops.py", line 88, in _nanargmax
    raise TypeError(
TypeError: reduction operation 'argmax' not allowed for this dtype
```

```
!python eval.py
```

```
2023-05-02 03:09:26.859363: I tensorflow/core/platform/cpu_feature_guard.cc:182] This TensorFlow binary is optimized
To enable the following instructions: AVX2 FMA, in other operations, rebuild TensorFlow with the appropriate compiler flags
2023-05-02 03:09:28.671324: W tensorflow/compiler/tf2tensorrt/utils/py_utils.cc:38] TF-TRT Warning: Could not find TensorRT
Evaluating roberta roberta-large v1
0% 2/1000 [00:01<10:01, 1.66it/s]
100% 20/20 [00:14<00:00, 1.38it/s]
```

```

/content/drive/.shortcut-targets-by-id/1xD1ZlQhDatgu8BZrlnav5pFJeOkI8nOi/depression-detection-lt-edi-2022/model
self.results = self.results.append(result, ignore_index=True)
Evaluating roberta roberta-large v2
0% 2/1000 [00:01<14:17, 1.16it/s]
100% 20/20 [00:14<00:00, 1.37it/s]
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1344: UndefinedMetricWarning: Precision
_warn_prf(average, modifier, msg_start, len(result))
/content/drive/.shortcut-targets-by-id/1xD1ZlQhDatgu8BZrlnav5pFJeOkI8nOi/depression-detection-lt-edi-2022/model
self.results = self.results.append(result, ignore_index=True)
roberta-large & 0.706 & 0.839 & 0.51 & 0.503 \\
Evaluating bert bert-base-multilingual-cased v1
0% 2/1000 [00:01<12:10, 1.37it/s]
100% 20/20 [00:05<00:00, 3.84it/s]
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1344: UndefinedMetricWarning: Precision
_warn_prf(average, modifier, msg_start, len(result))
/content/drive/.shortcut-targets-by-id/1xD1ZlQhDatgu8BZrlnav5pFJeOkI8nOi/depression-detection-lt-edi-2022/model
self.results = self.results.append(result, ignore_index=True)
Evaluating bert bert-base-multilingual-cased v2
0% 2/1000 [00:01<14:38, 1.14it/s]
100% 20/20 [00:05<00:00, 3.81it/s]
/content/drive/.shortcut-targets-by-id/1xD1ZlQhDatgu8BZrlnav5pFJeOkI8nOi/depression-detection-lt-edi-2022/model
self.results = self.results.append(result, ignore_index=True)
bert-base-multilingual-cased & 0.672 & 0.659 & 0.569 & 0.579 \\
Traceback (most recent call last):
  File "/content/drive/.shortcut-targets-by-id/1xD1ZlQhDatgu8BZrlnav5pFJeOkI8nOi/depression-detection-lt-edi-2(
    print(agg.get_result())
  File "/content/drive/.shortcut-targets-by-id/1xD1ZlQhDatgu8BZrlnav5pFJeOkI8nOi/depression-detection-lt-edi-2(
    values = self.get_best_result_by_main_metric(main_metric)
  File "/content/drive/.shortcut-targets-by-id/1xD1ZlQhDatgu8BZrlnav5pFJeOkI8nOi/depression-detection-lt-edi-2(
    return self.results.iloc[self.results[main_metric].idxmax()]
  File "/usr/local/lib/python3.10/dist-packages/pandas/core/series.py", line 2564, in idxmax
    i = self.argmax(axis, skipna, *args, **kwargs)
  File "/usr/local/lib/python3.10/dist-packages/pandas/core/base.py", line 655, in argmax
    return nanops.nanargmax( # type: ignore[return-value]
  File "/usr/local/lib/python3.10/dist-packages/pandas/core/nanops.py", line 88, in _f
    raise TypeError(
TypeError: reduction operation 'argmax' not allowed for this dtype

```

```
!python predict.py
```

```

2023-05-02 03:14:26.433727: I tensorflow/core/platform/cpu_feature_guard.cc:182] This TensorFlow binary is opti
To enable the following instructions: AVX2 FMA, in other operations, rebuild TensorFlow with the appropriate cc
2023-05-02 03:14:28.182058: W tensorflow/compiler/tf2tensorrt/utils/py_utils.cc:38] TF-TRT Warning: Could not f
Generating predictions using roberta rafalposwiata/roberta-large-depression None
Downloading (...)lve/main/config.json: 100% 904/904 [00:00<00:00, 4.37MB/s]
Downloading pytorch_model.bin: 100% 1.42G/1.42G [00:16<00:00, 84.2MB/s]
Downloading (...)okenizer_config.json: 100% 352/352 [00:00<00:00, 1.59MB/s]
Downloading (...)olve/main/vocab.json: 100% 798k/798k [00:00<00:00, 8.99MB/s]
Downloading (...)olve/main/merges.txt: 100% 456k/456k [00:00<00:00, 90.0MB/s]
Downloading (...)main/tokenizer.json: 100% 1.36M/1.36M [00:00<00:00, 49.6MB/s]
Downloading (...)cial_tokens_map.json: 100% 239/239 [00:00<00:00, 1.39MB/s]
0% 7/3245 [00:03<25:53, 2.08it/s]
100% 406/406 [00:21<00:00, 18.69it/s]
Generating predictions using roberta rafalposwiata/deproberta-large-depression None
Downloading (...)lve/main/config.json: 100% 924/924 [00:00<00:00, 5.73MB/s]
Downloading pytorch_model.bin: 100% 1.42G/1.42G [00:22<00:00, 63.9MB/s]
Downloading (...)okenizer_config.json: 100% 1.17k/1.17k [00:00<00:00, 7.41MB/s]
Downloading (...)olve/main/vocab.json: 100% 798k/798k [00:00<00:00, 201MB/s]
Downloading (...)olve/main/merges.txt: 100% 456k/456k [00:00<00:00, 316MB/s]
Downloading (...)main/tokenizer.json: 100% 1.36M/1.36M [00:00<00:00, 51.2MB/s]
Downloading (...)cial_tokens_map.json: 100% 772/772 [00:00<00:00, 4.36MB/s]
0% 7/3245 [00:03<27:30, 1.96it/s]
100% 406/406 [00:22<00:00, 18.40it/s]
0% 7/3245 [00:04<31:37, 1.71it/s]
100% 406/406 [00:22<00:00, 18.32it/s]
0% 7/3245 [00:03<29:40, 1.82it/s]
100% 406/406 [00:22<00:00, 18.16it/s]

```

```
!python reddit_depression_corpora.py
```

```

100% 81/81 [01:44<00:00, 1.29s/it]
depression: 94039 (23.8%)
anxiety: 53788 (13.6%)
suicidewatch: 34737 (8.8%)
lonely: 20095 (5.1%)
socialanxiety: 19652 (5.0%)

```

```

adhd: 10000 (2.5%)
conspiracy: 10000 (2.5%)
divorce: 10000 (2.5%)
EDAnonymous: 10000 (2.5%)
fitness: 10000 (2.5%)
guns: 10000 (2.5%)
jokes: 10000 (2.5%)
legaladvice: 10000 (2.5%)
mentalhealth: 10000 (2.5%)
parenting: 10000 (2.5%)
personalfinance: 10000 (2.5%)
relationships: 10000 (2.5%)
meditation: 7851 (2.0%)
autism: 7753 (2.0%)
bpd: 7304 (1.8%)
schizophrenia: 6656 (1.7%)
healthanxiety: 5428 (1.4%)
bipolarreddit: 5193 (1.3%)
alcoholism: 4628 (1.2%)
teaching: 3515 (0.9%)
ptsd: 2093 (0.5%)
addiction: 1758 (0.4%)
COVID19: 978 (0.2%)

```

Unique texts: 395468 (train: 98000 / validation: 2000).

```
!pip install -U deep-translator
```

Looking in indexes: <https://pypi.org/simple>, <https://us-python.pkg.dev/colab-wheels/public/simple/>

Collecting deep-translator

Downloading deep_translator-1.10.1-py3-none-any.whl (35 kB)

Requirement already satisfied: beautifulsoup4<5.0.0,>=4.9.1 in /usr/local/lib/python3.9/dist-packages (from deep-translator)

Requirement already satisfied: requests<3.0.0,>=2.23.0 in /usr/local/lib/python3.9/dist-packages (from deep-translator)

Requirement already satisfied: soupsieve>1.2 in /usr/local/lib/python3.9/dist-packages (from beautifulsoup4<5.0.0,>=4.9.1)

Requirement already satisfied: charset-normalizer~=2.0.0 in /usr/local/lib/python3.9/dist-packages (from requests<3.0.0,>=2.23.0)

Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.9/dist-packages (from requests<3.0.0,>=2.23.0)

Requirement already satisfied: urllib3<1.27,>=1.21.1 in /usr/local/lib/python3.9/dist-packages (from requests<3.0.0,>=2.23.0)

Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.9/dist-packages (from requests<3.0.0,>=2.23.0)

Installing collected packages: deep-translator

Successfully installed deep-translator-1.10.1

```
import pandas as pd
```

```
from deep_translator import GoogleTranslator
```

```
# Define languages to translate to
```

```
languages = ['en', 'de', 'es', 'hi']
```

```
# Read TSV file into DataFrame
```

```
df = pd.read_csv('/content/drive/MyDrive/depression-detection-lt-edi-2022/data/preprocessed_dataset/train_1.csv', sep='\t')
print(df)
```

```
# Iterate through each language and translate non-English sentences to that language
```

```
for lang in languages:
```

```
    # Create a new column in the DataFrame for the translated sentences
```

```
    # Iterate through each row and translate non-English sentences
```

```
    for index, row in df.iterrows():
```

```
        print(index, lang)
```

```
        if (len(row['text'])) < 5000:
```

```
            translated = GoogleTranslator(source='auto', target=lang).translate(row['text'])
```

```
            #translated = translator.translate(row['sentence']).text
```

```
            df.loc[index, 'translated_text'] = translated
```

```
            df.loc[index, 'language_ch'] = lang
```

```
    print(lang)
```

```
    # Append to file
```

```
    with open('/content/drive/MyDrive/depression-detection-lt-edi-2022/data/preprocessed_dataset/train_multi_lang.csv', 'a') as f:
```

```
        df.to_csv(f, sep='\t', index=False, header=(lang=='en'), mode='a')
```

Streaming output truncated to the last 5000 lines.

```
1003 de
```

```
1004 de
```

```
1005 de
```

```
1006 de
```

```

1007 de
1008 de
1009 de
1010 de
1011 de
1012 de
1013 de
1014 de
1015 de
1016 de
1017 de
1018 de
1019 de
1020 de
1021 de
1022 de
1023 de
1024 de
1025 de
1026 de
1027 de
1028 de
1029 de
1030 de
1031 de
1032 de
1033 de
1034 de
1035 de
1036 de
1037 de
1038 de
1039 de
1040 de
1041 de
1042 de
1043 de
1044 de
1045 de
1046 de
1047 de
1048 de
1049 de
1050 de
1051 de
1052 de
1053 de
1054 de
1055 de
1056 de
1057 de
1058 de

```

```
!python deproberta.py
```

```

2023-04-06 04:29:17.765418: I tensorflow/core/platform/cpu_feature_guard.cc:182] This TensorFlow binary is optimized
To enable the following instructions: AVX2 FMA, in other operations, rebuild TensorFlow with the appropriate compiler flags
2023-04-06 04:29:19.079574: W tensorflow/compiler/tf2tensorrt/utils/py_utils.cc:38] TF-TRT Warning: Could not find TensorRT
INFO:simpletransformers.language_modeling.language_modeling_utils: Creating features from dataset file at train_data.csv
100% 263924/263924 [05:53<00:00, 747.60it/s]
INFO:simpletransformers.language_modeling.language_modeling_utils: Saving features into cached file trained_model.pkl
/usr/local/lib/python3.9/dist-packages/transformers/optimization.py:391: FutureWarning: This implementation of AdamW is deprecated. Please use the
warnings.warn(
INFO:simpletransformers.language_modeling.language_modeling_model: Training started
Epoch 1 of 10: 0% 0/10 [00:00<?, ?it/s]
Running Epoch 0 of 10: 0% 0/13516 [00:00<?, ?it/s]
Epochs 0/10. Running Loss: 1.5115: 0% 0/13516 [00:01<?, ?it/s]/usr/local/lib/python3.9/dist-packages/torch/optim/adamw.py:11: FutureWarning: 'warn' is deprecated in favor of 'warning'
warnings.warn("Detected call of `lr_scheduler.step()` before `optimizer.step()`. ")

Epochs 0/10. Running Loss: 1.5115: 0% 1/13516 [00:02<7:38:22, 2.03s/it]
Epochs 0/10. Running Loss: 1.7326: 0% 1/13516 [00:02<9:42:43, 2.59s/it]
Epoch 1 of 10: 0% 0/10 [00:02<?, ?it/s]
Traceback (most recent call last):
  File "/content/drive/MyDrive/Colab Notebooks/Project_NLP/depression-detection-lt-edi-2022/models/deproberta.py", line 10, in <module>
    deproberta.train()
  File "/content/drive/MyDrive/Colab Notebooks/Project_NLP/depression-detection-lt-edi-2022/models/deproberta.py", line 10, in deproberta.train()
    self.model.train_model(train_data, eval_file=val_data)

```

```

File "/usr/local/lib/python3.9/dist-packages/simpletransformers/language_modeling/language_modeling_model.py"
  global_step, training_details = self.train(
File "/usr/local/lib/python3.9/dist-packages/simpletransformers/language_modeling/language_modeling_model.py"
  scaler.scale(loss).backward()
File "/usr/local/lib/python3.9/dist-packages/torch/_tensor.py", line 487, in backward
  torch.autograd.backward(
File "/usr/local/lib/python3.9/dist-packages/torch/autograd/__init__.py", line 200, in backward
  Variable._execution_engine.run_backward( # Calls into the C++ engine to run the backward pass
torch.cuda.OutOfMemoryError: CUDA out of memory. Tried to allocate 1.20 GiB (GPU 0; 14.75 GiB total capacity; 1

# Define a dictionary to map labels to label IDs
label_id_map = {
    0: "severe",
    1: "moderate",
    2: "not depression"
}

import pandas as pd

# Read the TSV file into a pandas DataFrame
df = pd.read_csv("/content/drive/MyDrive/depression-detection-lt-edi-2022/data/original_dataset/train_1.tsv", sep="\t")

# Update the column header "text" to "sentence"
#df = df.rename(columns={"Text": "sentence"})
#df = df.rename(columns={"final_label": "label_name"})
df['Label'] = df['Label'].map(label_id_map)
print(df.info())

# Save the updated DataFrame to a new TSV file
df.to_csv("/content/drive/MyDrive/depression-detection-lt-edi-2022/data/original_dataset/train_multi.tsv", sep="\t",

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8000 entries, 0 to 7999
Data columns (total 3 columns):
#   Column  Non-Null Count  Dtype
---  -
0    pid      8000 non-null    object
1    text      7892 non-null    object
2    Label     8000 non-null    object
dtypes: object(3)
memory usage: 187.6+ KB
None

import pandas as pd

# read the TSV file into a pandas dataframe
df = pd.read_csv('/content/drive/MyDrive/depression-detection-lt-edi-2022/data/original_dataset/train_multi.tsv', sep=

# remove rows where the text column is empty
df = df[df['text'].notna()]

# write the modified dataframe back to TSV file
df.to_csv('/content/drive/MyDrive/depression-detection-lt-edi-2022/data/original_dataset/train_multi_clean.tsv', sep=

```


✓ 0s completed at 11:19 PM

● ✕