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

```
Mounted at /content/drive
```

```
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/
Collecting transformers
  Downloading transformers-4.28.1-py3-none-any.whl (7.0 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 7.0/7.0 MB 96.0 MB/s eta 0:00:00
Collecting simpletransformers
  Downloading simpletransformers-0.63.11-py3-none-any.whl (250 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 250.7/250.7 kB 29.5 MB/s eta 0:00:00
Collecting pandas==1.2.5
  Downloading pandas-1.2.5.tar.gz (5.5 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 5.5/5.5 MB 17.5 MB/s eta 0:00:00
  Installing build dependencies ... done
  Getting requirements to build wheel ... done
  Preparing metadata (pyproject.toml) ... done
Collecting scikit-learn==0.23.1
  Downloading scikit-learn-0.23.1.tar.gz (7.2 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 7.2/7.2 MB 18.8 MB/s eta 0:00:00
  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/
Collecting transformers
  Using cached transformers-4.28.1-py3-none-any.whl (7.0 MB)
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-packages (from transformers) (3.12.0)
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.10/dist-packages (from transformers) (6.0)
Collecting huggingface-hub<1.0,>=0.11.0
  Downloading huggingface_hub-0.14.1-py3-none-any.whl (224 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 224.5/224.5 kB 20.4 MB/s eta 0:00:00
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.10/dist-packages (from transformers) (1.22.4)
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.10/dist-packages (from transformers) (2022.10.31)
Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-packages (from transformers) (2.27.1)
Requirement already satisfied: tqdm>=4.27 in /usr/local/lib/python3.10/dist-packages (from transformers) (4.65.0)
Collecting tokenizers!=0.11.3,<0.14,>=0.11.1
  Downloading tokenizers-0.13.3-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (7.8 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 7.8/7.8 MB 103.6 MB/s eta 0:00:00
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from transformers) (23.1)
Requirement already satisfied: fsspec in /usr/local/lib/python3.10/dist-packages (from huggingface-hub<1.0,>=0.11.0->transfo
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.10/dist-packages (from huggingface-hub<1.
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests->transformers) (3.4)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests->transformers) (2
Requirement already satisfied: charset-normalizer~=2.0.0 in /usr/local/lib/python3.10/dist-packages (from requests->transfor
Requirement already satisfied: urllib3<1.27,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests->transformers)
Installing collected packages: tokenizers, huggingface-hub, transformers
Successfully installed huggingface-hub-0.14.1 tokenizers-0.13.3 transformers-4.28.1
```

```
!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)
Collecting datasets
  Downloading datasets-2.12.0-py3-none-any.whl (474 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 474.6/474.6 kB 24.7 MB/s eta 0:00:00
Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-packages (from simpletransformers) (2.27.1)
Requirement already satisfied: scipy in /usr/local/lib/python3.10/dist-packages (from simpletransformers) (1.10.1)
Requirement already satisfied: scikit-learn in /usr/local/lib/python3.10/dist-packages (from simpletransformers) (1.2.2)
Collecting sentencepiece
  Downloading sentencepiece-0.1.99-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (1.3 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 1.3/1.3 MB 59.7 MB/s eta 0:00:00
Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages (from simpletransformers) (1.5.3)
Requirement already satisfied: numpy in /usr/local/lib/python3.10/dist-packages (from simpletransformers) (1.22.4)
Requirement already satisfied: tensorboard in /usr/local/lib/python3.10/dist-packages (from simpletransformers) (2.12.2)
Requirement already satisfied: tokenizers in /usr/local/lib/python3.10/dist-packages (from simpletransformers) (0.13.3)
Requirement already satisfied: tqdm>=4.47.0 in /usr/local/lib/python3.10/dist-packages (from simpletransformers) (4.65.0)
Collecting streamlit
  Downloading streamlit-1.22.0-py2.py3-none-any.whl (8.9 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 8.9/8.9 MB 78.5 MB/s eta 0:00:00
Requirement already satisfied: transformers>=4.6.0 in /usr/local/lib/python3.10/dist-packages (from simpletransformers) (4.28
Collecting seqeval
  Downloading seqeval-1.2.2.tar.gz (43 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 43.6/43.6 kB 5.4 MB/s eta 0:00:00
  Preparing metadata (setup.py) ... done
Requirement already satisfied: regex in /usr/local/lib/python3.10/dist-packages (from simpletransformers) (2022.10.31)
Collecting wandb>=0.10.32
  Downloading wandb-0.15.2-py3-none-any.whl (2.0 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 2.0/2.0 MB 89.6 MB/s eta 0:00:00
Requirement already satisfied: huggingface-hub<1.0,>=0.11.0 in /usr/local/lib/python3.10/dist-packages (from transformers>=4.
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from transformers>=4.6.0->simpletr
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.10/dist-packages (from transformers>=4.6.0->simpletrans
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-packages (from transformers>=4.6.0->simpletransform
Collecting GitPython!=3.1.29,>=1.0.0
  Downloading GitPython-3.1.31-py3-none-any.whl (184 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 184.3/184.3 kB 24.4 MB/s eta 0:00:00
Requirement already satisfied: Click!=8.0.0,>=7.0 in /usr/local/lib/python3.10/dist-packages (from wandb>=0.10.32->simpletran
Requirement already satisfied: appdirs>=1.4.3 in /usr/local/lib/python3.10/dist-packages (from wandb>=0.10.32->simpletransfor
Requirement already satisfied: protobuf!=4.21.0,<5,>=3.19.0 in /usr/local/lib/python3.10/dist-packages (from wandb>=0.10.32->
Collecting sentry-sdk>=1.0.0
  Downloading sentry_sdk-1.22.1-py2.py3-none-any.whl (203 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 203.1/203.1 kB 29.3 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->simpletransform
Collecting docker-pycreds>=0.4.0
  Downloading docker_pycreds-0.4.0-py2.py3-none-any.whl (9.0 kB)
Collecting setproctitle
  Downloading setproctitle-1.3.2-cp310-cp310-manylinux_2_5_x86_64.manylinux1_x86_64.manylinux_2_17_x86_64.manylinux2014_x86_6
Collecting pathtools
  Downloading pathtools-0.1.2.tar.gz (11 kB)
  Preparing metadata (setup.py) ... done
Requirement already satisfied: setuptools in /usr/local/lib/python3.10/dist-packages (from wandb>=0.10.32->simpletransformers
Requirement already satisfied: charset-normalizer~=2.0.0 in /usr/local/lib/python3.10/dist-packages (from requests->simpletra
Requirement already satisfied: urllib3<1.27,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests->simpletransfo
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests->simpletransforme
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests->simpletransformers) (3
Collecting dill<0.3.7,>=0.3.0
  Downloading dill-0.3.6-py3-none-any.whl (110 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 110.5/110.5 kB 18.6 MB/s eta 0:00:00
```

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

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

Cleaning and creating the train and dev datasets

```
!python preprocess_dataset.py
```

```
Original train_multilingual_robustness_1
severe: 480
moderate: 2059
not depression: 1488
all: 4027
-------------------------------------
Original train_multilingual_robustness_1 - without duplicates
severe: 480
moderate: 2059
not depression: 1488
```

```
all: 4027
--------------------------------------
Original dev
severe: 360
moderate: 2306
not depression: 1830
all: 4496
--------------------------------------
Original dev - without duplicates
severe: 360
moderate: 2304
not depression: 1817
all: 4481
--------------------------------------
Train after preprocessing
severe: 470
moderate: 2009
not depression: 1448
all: 3927
--------------------------------------
```

```
!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/models
```

```
!python models_list.py
```

```
!python transformers_models.py
```

finetune the model

```
!python finetune.py
```

```
Epochs 0/2. Running Loss:    0.9848:  51% 126/246 [02:47<01:54,  1.05it/s]
Epochs 0/2. Running Loss:    0.9848:  52% 127/246 [02:47<01:53,  1.05it/s]
Epochs 0/2. Running Loss:    0.9698:  52% 127/246 [02:48<01:53,  1.05it/s]
Epochs 0/2. Running Loss:    0.9698:  52% 128/246 [02:48<01:52,  1.05it/s]
Epochs 0/2. Running Loss:    0.8148:  52% 128/246 [02:49<01:52,  1.05it/s]
Epochs 0/2. Running Loss:    0.8148:  52% 129/246 [02:49<01:51,  1.05it/s]
Epochs 0/2. Running Loss:    0.9195:  52% 129/246 [02:50<01:51,  1.05it/s]
Epochs 0/2. Running Loss:    0.9195:  53% 130/246 [02:50<01:51,  1.04it/s]
Epochs 0/2. Running Loss:    1.0856:  53% 130/246 [02:51<01:51,  1.04it/s]
Epochs 0/2. Running Loss:    1.0856:  53% 131/246 [02:51<01:50,  1.04it/s]
Epochs 0/2. Running Loss:    0.8379:  53% 131/246 [02:52<01:50,  1.04it/s]
Epochs 0/2. Running Loss:    0.8379:  54% 132/246 [02:52<01:49,  1.04it/s]
Epochs 0/2. Running Loss:    0.9368:  54% 132/246 [02:53<01:49,  1.04it/s]
Epochs 0/2. Running Loss:    0.9368:  54% 133/246 [02:53<01:48,  1.04it/s]
Epochs 0/2. Running Loss:    0.8885:  54% 133/246 [02:54<01:48,  1.04it/s]
Epochs 0/2. Running Loss:    0.8885:  54% 134/246 [02:54<01:47,  1.04it/s]
Epochs 0/2. Running Loss:    1.0932:  54% 134/246 [02:55<01:47,  1.04it/s]
Epochs 0/2. Running Loss:    1.0932:  55% 135/246 [02:55<01:46,  1.04it/s]
Epochs 0/2. Running Loss:    1.0148:  55% 135/246 [02:56<01:46,  1.04it/s]
Epochs 0/2. Running Loss:    1.0148:  55% 136/246 [02:56<01:45,  1.05it/s]
Epochs 0/2. Running Loss:    0.9833:  55% 136/246 [02:57<01:45,  1.05it/s]
Epochs 0/2. Running Loss:    0.9833:  56% 137/246 [02:57<01:43,  1.05it/s]
Epochs 0/2. Running Loss:    0.7502:  56% 137/246 [02:57<01:43,  1.05it/s]
Epochs 0/2. Running Loss:    0.7502:  56% 138/246 [02:58<01:43,  1.05it/s]
Epochs 0/2. Running Loss:    0.9652:  56% 138/246 [02:58<01:43,  1.05it/s]
Epochs 0/2. Running Loss:    0.9652:  57% 139/246 [02:59<01:42,  1.05it/s]
Epochs 0/2. Running Loss:    0.7435:  57% 139/246 [02:59<01:42,  1.05it/s]
Epochs 0/2. Running Loss:    0.7435:  57% 140/246 [03:00<01:41,  1.05it/s]
Epochs 0/2. Running Loss:    1.0466:  57% 140/246 [03:00<01:41,  1.05it/s]
Epochs 0/2. Running Loss:    1.0466:  57% 141/246 [03:01<01:39,  1.05it/s]
Epochs 0/2. Running Loss:    1.0439:  57% 141/246 [03:01<01:39,  1.05it/s]
Epochs 0/2. Running Loss:    1.0439:  58% 142/246 [03:02<01:38,  1.05it/s]
Epochs 0/2. Running Loss:    0.9301:  58% 142/246 [03:02<01:38,  1.05it/s]
Epochs 0/2. Running Loss:    0.9301:  58% 143/246 [03:03<01:38,  1.05it/s]
Epochs 0/2. Running Loss:    0.9527:  58% 143/246 [03:03<01:38,  1.05it/s]
Epochs 0/2. Running Loss:    0.9527:  59% 144/246 [03:04<01:37,  1.05it/s]
Epochs 0/2. Running Loss:    1.0381:  59% 144/246 [03:04<01:37,  1.05it/s]
Epochs 0/2. Running Loss:    1.0381:  59% 145/246 [03:05<01:36,  1.05it/s]
Epochs 0/2. Running Loss:    1.0552:  59% 145/246 [03:05<01:36,  1.05it/s]
Epochs 0/2. Running Loss:    1.0552:  59% 146/246 [03:06<01:35,  1.05it/s]
Epochs 0/2. Running Loss:    0.8549:  59% 146/246 [03:06<01:35,  1.05it/s]
Epochs 0/2. Running Loss:    0.8549:  60% 147/246 [03:07<01:34,  1.05it/s]
Epochs 0/2. Running Loss:    0.8864:  60% 147/246 [03:07<01:34,  1.05it/s]
Epochs 0/2. Running Loss:    0.8864:  60% 148/246 [03:08<01:33,  1.05it/s]
Epochs 0/2. Running Loss:    0.9921:  60% 148/246 [03:08<01:33,  1.05it/s]
```

Results of each model on the English dev set of 1000 sentences and Fine-tuned on Multilingual train set of around 4000 sentences with noisy data

```
!python eval.py
```

```
2023-05-08 03:57:52.032281: I tensorflow/core/platform/cpu_feature_guard.cc:182] This TensorFlow binary is optimized to use a
To enable the following instructions: AVX2 FMA, in other operations, rebuild TensorFlow with the appropriate compiler flags.
2023-05-08 03:57:53.538677: 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<12:42,  1.31it/s]
100% 20/20 [00:14<00:00,  1.35it/s]
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1344: UndefinedMetricWarning: Precision is ill-de
  _warn_prf(average, modifier, msg_start, len(result))
/content/drive/.shortcut-targets-by-id/1xD1ZIqHdatgu8BZr1nav5pFJeOkI8nOi/depression-detection-lt-edi-2022/models/eval.py:20:
  self.results = self.results.append(result, ignore_index=True)
Evaluating roberta      roberta-large   v2
  0% 2/1000 [00:01<09:19,  1.78it/s]
100% 20/20 [00:14<00:00,  1.37it/s]
/content/drive/.shortcut-targets-by-id/1xD1ZIqHdatgu8BZr1nav5pFJeOkI8nOi/depression-detection-lt-edi-2022/models/eval.py:20:
  self.results = self.results.append(result, ignore_index=True)
roberta-large & 0.701 & 0.656 & 0.519 & 0.523 \\
Evaluating bert bert-base-multilingual-cased    v1
  0% 2/1000 [00:01<14:16,  1.16it/s]
100% 20/20 [00:05<00:00,  3.82it/s]
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1344: UndefinedMetricWarning: Precision is ill-de
  _warn_prf(average, modifier, msg_start, len(result))
/content/drive/.shortcut-targets-by-id/1xD1ZIqHdatgu8BZr1nav5pFJeOkI8nOi/depression-detection-lt-edi-2022/models/eval.py:20:
  self.results = self.results.append(result, ignore_index=True)
Evaluating bert bert-base-multilingual-cased    v2
  0% 2/1000 [00:01<11:13,  1.48it/s]
100% 20/20 [00:05<00:00,  3.80it/s]
/content/drive/.shortcut-targets-by-id/1xD1ZIqHdatgu8BZr1nav5pFJeOkI8nOi/depression-detection-lt-edi-2022/models/eval.py:20:
  self.results = self.results.append(result, ignore_index=True)
bert-base-multilingual-cased & 0.663 & 0.653 & 0.605 & 0.594 \\
```

```
Traceback (most recent call last):
  File "/content/drive/.shortcut-targets-by-id/1xD1ZIqHdatgu8BZr1nav5pFJeOkI8nOi/depression-detection-lt-edi-2022/models/eval
    result = agg.get_result()
  File "/content/drive/.shortcut-targets-by-id/1xD1ZIqHdatgu8BZr1nav5pFJeOkI8nOi/depression-detection-lt-edi-2022/models/eval
    values = self._get_best_result_by_main_metric(main_metric)
  File "/content/drive/.shortcut-targets-by-id/1xD1ZIqHdatgu8BZr1nav5pFJeOkI8nOi/depression-detection-lt-edi-2022/models/eval
    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
```

Results of each model on the English dev set of 1000 sentences and Fine-tuned on Multilingual train set of around 4000 sentences

```
!python eval.py
```

```
2023-05-02 22:39:17.561911: I tensorflow/core/platform/cpu_feature_guard.cc:182] This TensorFlow binary is optimized to use a
To enable the following instructions: AVX2 FMA, in other operations, rebuild TensorFlow with the appropriate compiler flags.
2023-05-02 22:39:18.898894: 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<10:09,  1.64it/s]
100% 20/20 [00:14<00:00,  1.35it/s]
/content/drive/.shortcut-targets-by-id/1xD1ZIqHdatgu8BZr1nav5pFJeOkI8nOi/depression-detection-lt-edi-2022/models/eval.py:20:
  self.results = self.results.append(result, ignore_index=True)
Evaluating roberta      roberta-large   v2
  0% 2/1000 [00:01<13:07,  1.27it/s]
100% 20/20 [00:14<00:00,  1.38it/s]
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1344: UndefinedMetricWarning: Precision is ill-de
  _warn_prf(average, modifier, msg_start, len(result))
/content/drive/.shortcut-targets-by-id/1xD1ZIqHdatgu8BZr1nav5pFJeOkI8nOi/depression-detection-lt-edi-2022/models/eval.py:20:
  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<11:23,  1.46it/s]
100% 20/20 [00:05<00:00,  3.88it/s]
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1344: UndefinedMetricWarning: Precision is ill-de
  _warn_prf(average, modifier, msg_start, len(result))
/content/drive/.shortcut-targets-by-id/1xD1ZIqHdatgu8BZr1nav5pFJeOkI8nOi/depression-detection-lt-edi-2022/models/eval.py:20:
  self.results = self.results.append(result, ignore_index=True)
Evaluating bert bert-base-multilingual-cased    v2
  0% 2/1000 [00:01<12:43,  1.31it/s]
100% 20/20 [00:05<00:00,  3.89it/s]
/content/drive/.shortcut-targets-by-id/1xD1ZIqHdatgu8BZr1nav5pFJeOkI8nOi/depression-detection-lt-edi-2022/models/eval.py:20:
  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/1xD1ZIqHdatgu8BZr1nav5pFJeOkI8nOi/depression-detection-lt-edi-2022/models/eval
    print(agg.get_result())
  File "/content/drive/.shortcut-targets-by-id/1xD1ZIqHdatgu8BZr1nav5pFJeOkI8nOi/depression-detection-lt-edi-2022/models/eval
    values = self._get_best_result_by_main_metric(main_metric)
  File "/content/drive/.shortcut-targets-by-id/1xD1ZIqHdatgu8BZr1nav5pFJeOkI8nOi/depression-detection-lt-edi-2022/models/eval
    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 optimized to use a
To enable the following instructions: AVX2 FMA, in other operations, rebuild TensorFlow with the appropriate compiler flags.
2023-05-02 03:14:28.182058: W tensorflow/compiler/tf2tensorrt/utils/py_utils.cc:38] TF-TRT Warning: Could not find TensorRT
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).
```

Below sections translate the datasets, cleaning and creating the csv's and tsv's

```
!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) (2.27
Requirement already satisfied: soupsieve>1.2 in /usr/local/lib/python3.9/dist-packages (from beautifulsoup4<5.0.0,>=4.9.1->de
Requirement already satisfied: charset-normalizer~=2.0.0 in /usr/local/lib/python3.9/dist-packages (from requests<3.0.0,>=2.2
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.9/dist-packages (from requests<3.0.0,>=2.23.0->de
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->deep-tra
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=',', encodir
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')
```

```
1276 de
1277 de
1278 de
1279 de
1280 de
1281 de
1282 de
1283 de
1284 de
1285 de
1286 de
1287 de
1288 de
1289 de
1290 de
1291 de
1292 de
1293 de
1294 de
1295 de
1296 de
1297 de
1298 de
1299 de
1300 de
1301 de
1302 de
1303 de
1304 de
1305 de
1306 de
1307 de
1308 de
1309 de
1310 de
1311 de
1312 de
1313 de
1314 de
1315 de
1316 de
1317 de
1318 de
1319 de
1320 de
1321 de
1322 de
1323 de
1324 de
1325 de
1326 de
1327 de
1328 de
1329 de
1330 de
1331 de
1332 de
1333 de
1334 de
```

```
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/preprocessed_dataset/train_multilang.csv', sep='\

# en_or_de_column = df[df["pid","language_ch","labels"].isin(["en", "de"])]["language_ch"]
# print(en_or_de_column)
```

```
# # remove rows where the text column is empty
# df = df[df['text'].notna()]
# #df = df.rename(columns={"Text": "sentence"})
# #df = df.rename(columns={"final_label": "label_name"})

# print(df.info())

# # 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='\t', index
import pandas as pd

# Load the CSV file into a pandas DataFrame
df = pd.read_csv('/content/drive/MyDrive/depression-detection-lt-edi-2022/data/preprocessed_dataset/train_multilang.csv', sep='\t'

# Filter the DataFrame to only include rows where "language_ch" is "en" or "de"
filtered_df = df[df["language_ch"].isin(["en"])]

new_df = filtered_df[["pid", "translated_text", "labels"]]
df = new_df.rename(columns={"translated_text": "text"})
df = df.rename(columns={"labels": "Label"})
df = df[df['text'].notna()]

# Write the filtered DataFrame to a new CSV file
df.to_csv("/content/drive/MyDrive/depression-detection-lt-edi-2022/data/original_dataset/eng_de.tsv" ,sep='\t', index=False)


import pandas as pd
import numpy as np
# read the TSV file into a pandas dataframe
# df = pd.read_csv('/content/drive/MyDrive/depression-detection-lt-edi-2022/data/preprocessed_dataset/train_multilang.csv', sep='\

# en_or_de_column = df[df["pid","language_ch","labels"].isin(["en", "de"])]["language_ch"]
# print(en_or_de_column)
# # remove rows where the text column is empty
# df = df[df['text'].notna()]
# #df = df.rename(columns={"Text": "sentence"})
# #df = df.rename(columns={"final_label": "label_name"})

# print(df.info())

# # 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='\t', index
import pandas as pd

# Load the CSV file into a pandas DataFrame
df = pd.read_csv('/content/drive/MyDrive/depression-detection-lt-edi-2022/data/original_dataset/train_multilingual_robustness.tsv'

# Filter the DataFrame to only include rows where "language_ch" is "en" or "de"
# filtered_df = df[df["language_ch"].isin(["de"])]

# new_df = filtered_df[["pid", "translated_text", "labels"]]
# df = new_df.rename(columns={"translated_text": "text"})
df = df.rename(columns={"labels": "Label"})
df = df[df['text'].notna()]
# num_rows = len(dev)
# random_order = np.random.permutation(num_rows)
# shuffled_df = dev.iloc[random_order]
# dev = shuffled_df.sample(frac=1, replace=False).head(1000)
# Write the filtered DataFrame to a new CSV file
df.to_csv("/content/drive/MyDrive/depression-detection-lt-edi-2022/data/original_dataset/train_multilingual_robust.tsv" ,sep='\t',


import random
import pandas as pd

# Read the CSV files and concatenate them into one DataFrame
df1 = pd.read_csv('/content/drive/MyDrive/depression-detection-lt-edi-2022/data/original_dataset/en.tsv', sep='\t')
#df2 = pd.read_csv('/content/drive/MyDrive/depression-detection-lt-edi-2022/data/original_dataset/de.tsv', sep='\t')
df3 = pd.read_csv('/content/drive/MyDrive/depression-detection-lt-edi-2022/data/original_dataset/es.tsv', sep='\t')
df4 = pd.read_csv('/content/drive/MyDrive/depression-detection-lt-edi-2022/data/original_dataset/hi.tsv', sep='\t')

df = pd.concat([df1, df3, df4], ignore_index=True)

# Choose 500 random rows from the DataFrame
random_indices = random.sample(range(len(df)), 2000)
random_df = df.iloc[random_indices]

# Append the selected rows to a new CSV file
```

```
#random_df.to_csv('/content/drive/MyDrive/depression-detection-lt-edi-2022/data/original_dataset/dev_out.tsv', 'a', index=False, h
random_df.to_csv('/content/drive/MyDrive/depression-detection-lt-edi-2022/data/original_dataset/dev_out.tsv', mode='a', index=Fals
```

```
!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 use a
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 trained_models/depr
100% 263924/263924 [05:53<00:00, 747.60it/s]
INFO:simpletransformers.language_modeling.language_modeling_utils: Saving features into cached file trained_models/deproberta
/usr/local/lib/python3.9/dist-packages/transformers/optimization.py:391: FutureWarning: This implementation of AdamW is depre
  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/lr_sche
  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 52, i
    deproberta.train()
  File "/content/drive/MyDrive/Colab Notebooks/Project_NLP/depression-detection-lt-edi-2022/models/deproberta.py", line 34, i
    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", line 463, in
    global_step, training_details = self.train(
  File "/usr/local/lib/python3.9/dist-packages/simpletransformers/language_modeling/language_modeling_model.py", line 854, in
    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; 12.79 GiB alrea
```

```
# 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/preprocessed_dataset/train_multilingual_robustness.
print(df.columns)
# Update the column header "text" to "sentence"
df = df.rename(columns={"labels": "Label"})
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/preprocessed_dataset/train_multilingual_robustness_1.csv",
```

```
Index(['pid', 'text', 'labels'], dtype='object')
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4027 entries, 0 to 4026
Data columns (total 3 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   pid     4027 non-null   object
 1   text    4027 non-null   object
 2   Label   4027 non-null   object
dtypes: object(3)
memory usage: 94.5+ 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='\t')

# 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='\t', index=F
```