

# Base\_BERT\_Disaster\_Tweets

May 16, 2022

```
[ ]: !pip install datasets  
      !pip install transformers
```

```
Requirement already satisfied: datasets in /usr/local/lib/python3.7/dist-  
packages (2.2.1)  
Requirement already satisfied: xxhash in /usr/local/lib/python3.7/dist-packages  
(from datasets) (3.0.0)  
Requirement already satisfied: fsspec[http]>=2021.05.0 in  
/usr/local/lib/python3.7/dist-packages (from datasets) (2022.3.0)  
Requirement already satisfied: requests>=2.19.0 in  
/usr/local/lib/python3.7/dist-packages (from datasets) (2.23.0)  
Requirement already satisfied: pyarrow>=6.0.0 in /usr/local/lib/python3.7/dist-  
packages (from datasets) (6.0.1)  
Requirement already satisfied: dill in /usr/local/lib/python3.7/dist-packages  
(from datasets) (0.3.4)  
Requirement already satisfied: responses<0.19 in /usr/local/lib/python3.7/dist-  
packages (from datasets) (0.18.0)  
Requirement already satisfied: tqdm>=4.62.1 in /usr/local/lib/python3.7/dist-  
packages (from datasets) (4.64.0)  
Requirement already satisfied: multiprocessing in /usr/local/lib/python3.7/dist-  
packages (from datasets) (0.70.12.2)  
Requirement already satisfied: huggingface-hub<1.0.0,>=0.1.0 in  
/usr/local/lib/python3.7/dist-packages (from datasets) (0.6.0)  
Requirement already satisfied: pandas in /usr/local/lib/python3.7/dist-packages  
(from datasets) (1.3.5)  
Requirement already satisfied: importlib-metadata in  
/usr/local/lib/python3.7/dist-packages (from datasets) (4.11.3)  
Requirement already satisfied: packaging in /usr/local/lib/python3.7/dist-  
packages (from datasets) (21.3)  
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.7/dist-  
packages (from datasets) (1.21.6)  
Requirement already satisfied: aiohttp in /usr/local/lib/python3.7/dist-packages  
(from datasets) (3.8.1)  
Requirement already satisfied: pyyaml in /usr/local/lib/python3.7/dist-packages  
(from huggingface-hub<1.0.0,>=0.1.0->datasets) (6.0)  
Requirement already satisfied: filelock in /usr/local/lib/python3.7/dist-  
packages (from huggingface-hub<1.0.0,>=0.1.0->datasets) (3.6.0)  
Requirement already satisfied: typing-extensions>=3.7.4.3 in
```

/usr/local/lib/python3.7/dist-packages (from huggingface-  
 hub<1.0.0,>=0.1.0->datasets) (4.2.0)  
 Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in  
 /usr/local/lib/python3.7/dist-packages (from packaging->datasets) (3.0.8)  
 Requirement already satisfied: urllib3!=1.25.0,!1.25.1,<1.26,>=1.21.1 in  
 /usr/local/lib/python3.7/dist-packages (from requests>=2.19.0->datasets)  
 (1.25.11)  
 Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-  
 packages (from requests>=2.19.0->datasets) (2.10)  
 Requirement already satisfied: certifi>=2017.4.17 in  
 /usr/local/lib/python3.7/dist-packages (from requests>=2.19.0->datasets)  
 (2021.10.8)  
 Requirement already satisfied: chardet<4,>=3.0.2 in  
 /usr/local/lib/python3.7/dist-packages (from requests>=2.19.0->datasets) (3.0.4)  
 Requirement already satisfied: multidict<7.0,>=4.5 in  
 /usr/local/lib/python3.7/dist-packages (from aiohttp->datasets) (6.0.2)  
 Requirement already satisfied: async-timeout<5.0,>=4.0.0a3 in  
 /usr/local/lib/python3.7/dist-packages (from aiohttp->datasets) (4.0.2)  
 Requirement already satisfied: aiosignal>=1.1.2 in  
 /usr/local/lib/python3.7/dist-packages (from aiohttp->datasets) (1.2.0)  
 Requirement already satisfied: async-test==0.13.0 in  
 /usr/local/lib/python3.7/dist-packages (from aiohttp->datasets) (0.13.0)  
 Requirement already satisfied: frozenlist>=1.1.1 in  
 /usr/local/lib/python3.7/dist-packages (from aiohttp->datasets) (1.3.0)  
 Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.7/dist-  
 packages (from aiohttp->datasets) (21.4.0)  
 Requirement already satisfied: yarll<2.0,>=1.0 in /usr/local/lib/python3.7/dist-  
 packages (from aiohttp->datasets) (1.7.2)  
 Requirement already satisfied: charset-normalizer<3.0,>=2.0 in  
 /usr/local/lib/python3.7/dist-packages (from aiohttp->datasets) (2.0.12)  
 Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.7/dist-  
 packages (from importlib-metadata->datasets) (3.8.0)  
 Requirement already satisfied: pytz>=2017.3 in /usr/local/lib/python3.7/dist-  
 packages (from pandas->datasets) (2022.1)  
 Requirement already satisfied: python-dateutil>=2.7.3 in  
 /usr/local/lib/python3.7/dist-packages (from pandas->datasets) (2.8.2)  
 Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/dist-  
 packages (from python-dateutil>=2.7.3->pandas->datasets) (1.15.0)  
 Requirement already satisfied: transformers in /usr/local/lib/python3.7/dist-  
 packages (4.19.1)  
 Requirement already satisfied: importlib-metadata in  
 /usr/local/lib/python3.7/dist-packages (from transformers) (4.11.3)  
 Requirement already satisfied: tqdm>=4.27 in /usr/local/lib/python3.7/dist-  
 packages (from transformers) (4.64.0)  
 Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.7/dist-  
 packages (from transformers) (1.21.6)  
 Requirement already satisfied: regex!=2019.12.17 in  
 /usr/local/lib/python3.7/dist-packages (from transformers) (2019.12.20)

Requirement already satisfied: requests in /usr/local/lib/python3.7/dist-packages (from transformers) (2.23.0)  
 Requirement already satisfied: tokenizers!=0.11.3,<0.13,>=0.11.1 in /usr/local/lib/python3.7/dist-packages (from transformers) (0.12.1)  
 Requirement already satisfied: huggingface-hub<1.0,>=0.1.0 in /usr/local/lib/python3.7/dist-packages (from transformers) (0.6.0)  
 Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.7/dist-packages (from transformers) (6.0)  
 Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.7/dist-packages (from transformers) (21.3)  
 Requirement already satisfied: filelock in /usr/local/lib/python3.7/dist-packages (from transformers) (3.6.0)  
 Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.7/dist-packages (from huggingface-hub<1.0,>=0.1.0->transformers) (4.2.0)  
 Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in /usr/local/lib/python3.7/dist-packages (from packaging>=20.0->transformers) (3.0.8)  
 Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.7/dist-packages (from importlib-metadata->transformers) (3.8.0)  
 Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/dist-packages (from requests->transformers) (3.0.4)  
 Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-packages (from requests->transformers) (2.10)  
 Requirement already satisfied: urllib3!=1.25.0,!1.25.1,<1.26,>=1.21.1 in /usr/local/lib/python3.7/dist-packages (from requests->transformers) (1.25.11)  
 Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-packages (from requests->transformers) (2021.10.8)

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

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force\_remount=True).

```
[ ]: import pandas as pd
      import numpy as np
      import torch
      from torch import cuda
      import random
      import os
      import torch
      from torch import nn
      from transformers import Trainer
      from collections import Counter
      from sklearn.model_selection import train_test_split
      from sklearn.metrics import f1_score, classification_report
      from sklearn.utils.class_weight import compute_class_weight
```

```

import datasets
from datasets import Dataset, load_metric
import transformers
from transformers import AutoTokenizer
from transformers import AutoModelForSequenceClassification, TrainingArguments,
    Trainer, DataCollatorWithPadding

print(torch.__version__)
print(transformers.__version__)

```

1.11.0+cu113

4.19.1

```
[ ]: cd /content/drive/MyDrive/Colab Notebooks
```

/content/drive/MyDrive/Colab Notebooks

```

[ ]: train_path = './nlp-getting-started/train.csv'
train_data = pd.read_csv(train_path)
train_data = train_data[['text', 'target']]
train_data.rename(columns={"target": "label"}, inplace=True)

test_path = './nlp-getting-started/train.csv'
test_data = pd.read_csv(test_path)
test_data = test_data[['text']]
test_data.rename(columns={"target": "label"}, inplace=True)

```

```

[ ]: train_data, dev_data = train_test_split(train_data, test_size=0.1,
    shuffle=True, stratify=train_data['label'])

print("Train dataset labels count = ", Counter(train_data['label']))
print("Dev dataset labels count = ", Counter(dev_data['label']))
#print("Test dataset labels count = ", Counter(test_data['target'])) #test
    dataset does not contain the target label

```

Train dataset labels count = Counter({0: 3907, 1: 2944})

Dev dataset labels count = Counter({0: 435, 1: 327})

```

[ ]: model_checkpoint = 'bert-base-uncased'
tokenizer = AutoTokenizer.from_pretrained(model_checkpoint, use_fast=False)
model = AutoModelForSequenceClassification.from_pretrained(model_checkpoint,
    num_labels=2).to('cuda')

```

loading configuration file <https://huggingface.co/bert-base-uncased/resolve/main/config.json> from cache at /root/.cache/huggingface/transformers/3c61d016573b14f7f008c02c4e51a366c67ab274726fe2910691e2a761acf43e.37395cee442ab11005bcd270f3c34464dc1704b715b5d7d52b1a461abe3b9e4e

Model config BertConfig {  
 "\_name\_or\_path": "bert-base-uncased",

```

"architectures": [
  "BertForMaskedLM"
],
"attention_probs_dropout_prob": 0.1,
"classifier_dropout": null,
"gradient_checkpointing": false,
"hidden_act": "gelu",
"hidden_dropout_prob": 0.1,
"hidden_size": 768,
"initializer_range": 0.02,
"intermediate_size": 3072,
"layer_norm_eps": 1e-12,
"max_position_embeddings": 512,
"model_type": "bert",
"num_attention_heads": 12,
"num_hidden_layers": 12,
"pad_token_id": 0,
"position_embedding_type": "absolute",
"transformers_version": "4.19.1",
"type_vocab_size": 2,
"use_cache": true,
"vocab_size": 30522
}

```

loading file <https://huggingface.co/bert-base-uncased/resolve/main/vocab.txt>  
from cache at /root/.cache/huggingface/transformers/45c3f7a79a80e1cf0a489e5c62b43f173c15db47864303a55d623bb3c96f72a5.d789d64ebfe299b0e416afc4a169632f903f693095b4629a7ea271d5a0cf2c99

loading file [https://huggingface.co/bert-base-uncased/resolve/main/added\\_tokens.json](https://huggingface.co/bert-base-uncased/resolve/main/added_tokens.json) from cache at None

loading file [https://huggingface.co/bert-base-uncased/resolve/main/special\\_tokens\\_map.json](https://huggingface.co/bert-base-uncased/resolve/main/special_tokens_map.json) from cache at None

loading file [https://huggingface.co/bert-base-uncased/resolve/main/tokenizer\\_config.json](https://huggingface.co/bert-base-uncased/resolve/main/tokenizer_config.json) from cache at /root/.cache/huggingface/transformers/c1d7f0a763fb63861cc08553866f1fc3e5a6f4f07621be277452d26d71303b7e.20430bd8e10ef77a7d2977accefe796051e01bc2fc4aa146bc862997a1a15e79

loading configuration file <https://huggingface.co/bert-base-uncased/resolve/main/config.json> from cache at /root/.cache/huggingface/transformers/3c61d016573b14f7f008c02c4e51a366c67ab274726fe2910691e2a761acf43e.37395cee442ab11005bcd270f3c34464dc1704b715b5d7d52b1a461abe3b9e4e

```

Model config BertConfig {
  "_name_or_path": "bert-base-uncased",
  "architectures": [
    "BertForMaskedLM"
  ],
  "attention_probs_dropout_prob": 0.1,
  "classifier_dropout": null,
  "gradient_checkpointing": false,

```

```

"hidden_act": "gelu",
"hidden_dropout_prob": 0.1,
"hidden_size": 768,
"initializer_range": 0.02,
"intermediate_size": 3072,
"layer_norm_eps": 1e-12,
"max_position_embeddings": 512,
"model_type": "bert",
"num_attention_heads": 12,
"num_hidden_layers": 12,
"pad_token_id": 0,
"position_embedding_type": "absolute",
"transformers_version": "4.19.1",
"type_vocab_size": 2,
"use_cache": true,
"vocab_size": 30522
}

```

loading configuration file <https://huggingface.co/bert-base-uncased/resolve/main/config.json> from cache at /root/.cache/huggingface/transformers/3c61d016573b14f7f008c02c4e51a366c67ab274726fe2910691e2a761acf43e.37395cee442ab11005bcd270f3c34464dc1704b715b5d7d52b1a461abe3b9e4e

```

Model config BertConfig {
  "name_or_path": "bert-base-uncased",
  "architectures": [
    "BertForMaskedLM"
  ],
  "attention_probs_dropout_prob": 0.1,
  "classifier_dropout": null,
  "gradient_checkpointing": false,
  "hidden_act": "gelu",
  "hidden_dropout_prob": 0.1,
  "hidden_size": 768,
  "initializer_range": 0.02,
  "intermediate_size": 3072,
  "layer_norm_eps": 1e-12,
  "max_position_embeddings": 512,
  "model_type": "bert",
  "num_attention_heads": 12,
  "num_hidden_layers": 12,
  "pad_token_id": 0,
  "position_embedding_type": "absolute",
  "transformers_version": "4.19.1",
  "type_vocab_size": 2,
  "use_cache": true,
  "vocab_size": 30522
}

```

loading weights file [https://huggingface.co/bert-base-uncased/resolve/main/pytorch\\_model.bin](https://huggingface.co/bert-base-uncased/resolve/main/pytorch_model.bin) from cache at /root/.cache/huggingface/transformers/a8041bf617d7f94ea26d15e218abd04afc2004805632abc0ed2066aa16d50d04.faf6ea826ae9c5867d12b22257f9877e6b8367890837bd60f7c54a29633f7f2f

Some weights of the model checkpoint at bert-base-uncased were not used when initializing BertForSequenceClassification: ['cls.seq\_relationship.bias', 'cls.predictions.transform.dense.bias', 'cls.predictions.transform.LayerNorm.bias', 'cls.seq\_relationship.weight', 'cls.predictions.transform.LayerNorm.weight', 'cls.predictions.decoder.weight', 'cls.predictions.transform.dense.weight', 'cls.predictions.bias']

- This IS expected if you are initializing BertForSequenceClassification 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 BertForSequenceClassification from the checkpoint of a model that you expect to be exactly identical (initializing a BertForSequenceClassification model from a BertForSequenceClassification model).

Some weights of BertForSequenceClassification were not initialized from the model checkpoint at bert-base-uncased and are newly initialized:

['classifier.bias', 'classifier.weight']

You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.

```
[ ]: def preprocess_function(examples):  
      return tokenizer(examples['text'], truncation=True)
```

```
[ ]: train_data = Dataset.from_pandas(train_data)  
      dev_data = Dataset.from_pandas(dev_data)  
      test_data = Dataset.from_pandas(test_data)  
  
      encoded_dataset_train = train_data.map(preprocess_function, batched=True)  
      encoded_dataset_dev = dev_data.map(preprocess_function, batched=True)  
      encoded_dataset_test = test_data.map(preprocess_function, batched=True)
```

```
0%|          | 0/7 [00:00<?, ?ba/s]
```

```
0%|          | 0/1 [00:00<?, ?ba/s]
```

```
0%|          | 0/8 [00:00<?, ?ba/s]
```

```
[ ]: columns_to_return = ['input_ids', 'label', 'attention_mask']  
      columns_to_return_test = ['input_ids', 'attention_mask']  
      encoded_dataset_train.set_format(columns=columns_to_return)  
      encoded_dataset_dev.set_format(columns=columns_to_return)  
      encoded_dataset_test.set_format(columns=columns_to_return_test)
```

```
[ ]: batch_size = 8  
      metric_name = "f1"
```

```

model_name = model_checkpoint.split("/")[-1]
task = 'tweet'

args = TrainingArguments(
    f"./save_model/{model_name}-finetuned-{task}",
    evaluation_strategy = "epoch",
    save_strategy = "epoch",
    learning_rate=1e-5,
    per_device_train_batch_size=batch_size,
    per_device_eval_batch_size=batch_size,
    num_train_epochs=4,
    weight_decay=0.01,
    load_best_model_at_end=True,
    metric_for_best_model=metric_name,
    push_to_hub=False,
)

```

PyTorch: setting up devices

The default value for the training argument `--report\_to` will change in v5 (from all installed integrations to none). In v5, you will need to use `--report\_to all` to get the same behavior as now. You should start updating your code and make this info disappear :-).

```

[ ]: metric = load_metric('f1')
def compute_metrics(eval_pred):
    predictions, labels = eval_pred
    predictions = np.argmax(predictions, axis=1)
    return metric.compute(predictions=predictions, references=labels,
↪average='macro')

```

```

[ ]: trainer = Trainer(
    model,
    args,
    train_dataset=encoded_dataset_train,
    eval_dataset=encoded_dataset_dev,
    tokenizer=tokenizer,
    compute_metrics=compute_metrics
)

```

```

[ ]: trainer.train()

```

The following columns in the training set don't have a corresponding argument in `BertForSequenceClassification.forward` and have been ignored: text, \_\_index\_level\_0\_\_. If text, \_\_index\_level\_0\_\_ are not expected by `BertForSequenceClassification.forward`, you can safely ignore this message.  
/usr/local/lib/python3.7/dist-packages/transformers/optimization.py:309:  
FutureWarning: This implementation of AdamW is deprecated and will be removed in a future version. Use the PyTorch implementation torch.optim.AdamW instead, or



```

set `no_deprecation_warning=True` to disable this warning
FutureWarning,
***** Running training *****
  Num examples = 6851
  Num Epochs = 4
  Instantaneous batch size per device = 8
  Total train batch size (w. parallel, distributed & accumulation) = 8
  Gradient Accumulation steps = 1
  Total optimization steps = 3428

<IPython.core.display.HTML object>

The following columns in the evaluation set don't have a corresponding argument
in `BertForSequenceClassification.forward` and have been ignored: text,
__index_level_0__. If text, __index_level_0__ are not expected by
`BertForSequenceClassification.forward`, you can safely ignore this message.
***** Running Evaluation *****
  Num examples = 762
  Batch size = 8
Saving model checkpoint to ./save_model/bert-base-uncased-finetuned-
tweet/checkpoint-857
Configuration saved in ./save_model/bert-base-uncased-finetuned-
tweet/checkpoint-857/config.json
Model weights saved in ./save_model/bert-base-uncased-finetuned-
tweet/checkpoint-857/pytorch_model.bin
tokenizer config file saved in ./save_model/bert-base-uncased-finetuned-
tweet/checkpoint-857/tokenizer_config.json
Special tokens file saved in ./save_model/bert-base-uncased-finetuned-
tweet/checkpoint-857/special_tokens_map.json
The following columns in the evaluation set don't have a corresponding argument
in `BertForSequenceClassification.forward` and have been ignored: text,
__index_level_0__. If text, __index_level_0__ are not expected by
`BertForSequenceClassification.forward`, you can safely ignore this message.
***** Running Evaluation *****
  Num examples = 762
  Batch size = 8
Saving model checkpoint to ./save_model/bert-base-uncased-finetuned-
tweet/checkpoint-1714
Configuration saved in ./save_model/bert-base-uncased-finetuned-
tweet/checkpoint-1714/config.json
Model weights saved in ./save_model/bert-base-uncased-finetuned-
tweet/checkpoint-1714/pytorch_model.bin
tokenizer config file saved in ./save_model/bert-base-uncased-finetuned-
tweet/checkpoint-1714/tokenizer_config.json
Special tokens file saved in ./save_model/bert-base-uncased-finetuned-
tweet/checkpoint-1714/special_tokens_map.json
The following columns in the evaluation set don't have a corresponding argument
in `BertForSequenceClassification.forward` and have been ignored: text,
__index_level_0__. If text, __index_level_0__ are not expected by

```

```

`BertForSequenceClassification.forward`, you can safely ignore this message.
***** Running Evaluation *****
  Num examples = 762
  Batch size = 8
Saving model checkpoint to ./save_model/bert-base-uncased-finetuned-
tweet/checkpoint-2571
Configuration saved in ./save_model/bert-base-uncased-finetuned-
tweet/checkpoint-2571/config.json
Model weights saved in ./save_model/bert-base-uncased-finetuned-
tweet/checkpoint-2571/pytorch_model.bin
tokenizer config file saved in ./save_model/bert-base-uncased-finetuned-
tweet/checkpoint-2571/tokenizer_config.json
Special tokens file saved in ./save_model/bert-base-uncased-finetuned-
tweet/checkpoint-2571/special_tokens_map.json
The following columns in the evaluation set don't have a corresponding argument
in `BertForSequenceClassification.forward` and have been ignored: text,
__index_level_0__. If text, __index_level_0__ are not expected by
`BertForSequenceClassification.forward`, you can safely ignore this message.
***** Running Evaluation *****
  Num examples = 762
  Batch size = 8
Saving model checkpoint to ./save_model/bert-base-uncased-finetuned-
tweet/checkpoint-3428
Configuration saved in ./save_model/bert-base-uncased-finetuned-
tweet/checkpoint-3428/config.json
Model weights saved in ./save_model/bert-base-uncased-finetuned-
tweet/checkpoint-3428/pytorch_model.bin
tokenizer config file saved in ./save_model/bert-base-uncased-finetuned-
tweet/checkpoint-3428/tokenizer_config.json
Special tokens file saved in ./save_model/bert-base-uncased-finetuned-
tweet/checkpoint-3428/special_tokens_map.json

Training completed. Do not forget to share your model on huggingface.co/models
=)

```

```

Loading best model from ./save_model/bert-base-uncased-finetuned-
tweet/checkpoint-1714 (score: 0.8339883748289776).

```

```

[ ]: TrainOutput(global_step=3428, training_loss=0.3453334421093294,
metrics={'train_runtime': 437.0492, 'train_samples_per_second': 62.702,
'train_steps_per_second': 7.844, 'total_flos': 709032391428420.0, 'train_loss':
0.3453334421093294, 'epoch': 4.0})

```

```

[ ]: #get_test_predictions
predictions = trainer.predict(encoded_dataset_test)

```

```

preds = np.argmax(predictions.predictions, axis=-1)
#print the top 100 examples
for i in range(100):
    print(encoded_dataset_test['text'][i], preds[i], sep='\t')

```

The following columns in the test set don't have a corresponding argument in `BertForSequenceClassification.forward` and have been ignored: text. If text are not expected by `BertForSequenceClassification.forward`, you can safely ignore this message.

\*\*\*\*\* Running Prediction \*\*\*\*\*

Num examples = 7613

Batch size = 8

<IPython.core.display.HTML object>

```

Our Deeds are the Reason of this #earthquake May ALLAH Forgive us all    1
Forest fire near La Ronge Sask. Canada    1
All residents asked to 'shelter in place' are being notified by officers. No
other evacuation or shelter in place orders are expected    1
13,000 people receive #wildfires evacuation orders in California    1
Just got sent this photo from Ruby #Alaska as smoke from #wildfires pours into a
school    1
#RockyFire Update => California Hwy. 20 closed in both directions due to Lake
County fire - #CAfire #wildfires    1
#flood #disaster Heavy rain causes flash flooding of streets in Manitou,
Colorado Springs areas    1
I'm on top of the hill and I can see a fire in the woods...    1
There's an emergency evacuation happening now in the building across the street
1
I'm afraid that the tornado is coming to our area...    1
Three people died from the heat wave so far    1
Haha South Tampa is getting flooded hah- WAIT A SECOND I LIVE IN SOUTH TAMPA
WHAT AM I GONNA DO WHAT AM I GONNA DO FVCK #flooding    1
#raining #flooding #Florida #TampaBay #Tampa 18 or 19 days. I've lost count
1
#Flood in Bago Myanmar #We arrived Bago    1
Damage to school bus on 80 in multi car crash #BREAKING    1
What's up man?    0
I love fruits    0
Summer is lovely    0
My car is so fast    0
What a goooooooooaaaaaal!!!!!!    0
this is ridiculous...    0
London is cool ;)    0
Love skiing    0
What a wonderful day!    0
LOOOOOOOL    0
No way...I can't eat that shit    0
Was in NYC last week!    0

```

Love my girlfriend 0  
 Coooool :) 0  
 Do you like pasta? 0  
 The end! 0  
 @bbcmtd Wholesale Markets ablaze <http://t.co/lHYXEOHY6C> 1  
 We always try to bring the heavy. #metal #RT <http://t.co/YAo1e0xngw> 0  
 #AFRICANBAZE: Breaking news:Nigeria flag set ablaze in Aba.  
<http://t.co/2nndBGwyEi> 1  
 Crying out for more! Set me ablaze 0  
 On plus side LOOK AT THE SKY LAST NIGHT IT WAS ABLAZE <http://t.co/qqsmsshaJ3N>  
 0  
 @PhDSquares #muftc they've built so much hype around new acquisitions but I doubt  
 they will set the EPL ablaze this season. 0  
 INEC Office in Abia Set Ablaze - <http://t.co/3ImaomknnA> 1  
 Barbados #Bridgetown JAMAICA Û Two cars set ablaze: SANTA CRUZ Û Head of the  
 St Elizabeth Police Superintende... <http://t.co/wDUEaj8Q4J> 1  
 Ablaze for you Lord :D 0  
 Check these out: <http://t.co/rOI2NSmEJJ> <http://t.co/3Tj8ZjiN21>  
<http://t.co/YDUiXefIpE> <http://t.co/LxTjc87KLS> #nsfw 0  
 on the outside you're ablaze and alive  
 but you're dead inside 0  
 Had an awesome time visiting the CFC head office the ancop site and ablaze.  
 Thanks to Tita Vida for taking care of us ?? 0  
 SOOOO PUMPED FOR ABLAZE ???? @southridgelife 0  
 I wanted to set Chicago ablaze with my preaching... But not my hotel!  
<http://t.co/o9qknbf0FX> 0  
 I gained 3 followers in the last week. You? Know your stats and grow with  
<http://t.co/TIyUliF5c6> 0  
 How the West was burned: Thousands of wildfires ablaze in California alone  
<http://t.co/vl5TBR3wbr> 1  
 Building the perfect tracklist to life leave the streets ablaze 0  
 Check these out: <http://t.co/rOI2NSmEJJ> <http://t.co/3Tj8ZjiN21>  
<http://t.co/YDUiXefIpE> <http://t.co/LxTjc87KLS> #nsfw 0  
 First night with retainers in. It's quite weird. Better get used to it; I have  
 to wear them every single night for the next year at least. 0  
 Deputies: Man shot before Brighton home set ablaze <http://t.co/gWNRhMSO8k>  
 1  
 Man wife get six years jail for setting ablaze niece  
<http://t.co/eV1ahOUCZA> 1  
 SANTA CRUZ Û Head of the St Elizabeth Police Superintendent Lanford Salmon has  
 r ... - <http://t.co/vplR5Hka2u> <http://t.co/SxHW2TNNLf> 0  
 Police: Arsonist Deliberately Set Black Church In North Carolinaã Ablaze  
<http://t.co/pcXarbH9An> 1  
 Noches El-Bestia ' @Alexis\_Sanchez: happy to see my teammates and training hard  
 ?? goodnight gunners.????? <http://t.co/uc4j4jHvGR> 0  
 #Kurds trampling on Turkmen flag later set it ablaze while others vandalized  
 offices of Turkmen Front in #Diyala <http://t.co/4IzFdYC3cg> 1  
 TRUCK ABLAZE : R21. VOORTREKKER AVE. OUTSIDE OR TAMBO INTL. CARGO SECTION.

<http://t.co/8kscqKfKkF> 1  
 Set our hearts ablaze and every city was a gift And every skyline was like a  
 kiss upon the lips @Ü\_ <https://t.co/cYoMPZ1AOZ> 0  
 They sky was ablaze tonight in Los Angeles. I'm expecting IG and FB to be filled  
 with sunset shots if I know my peeps!! 0  
 How the West was burned: Thousands of wildfires ablaze in #California alone  
<http://t.co/iCSjGZ9tE1> #climate #energy <http://t.co/9FxmN010Bd> 1  
 Revel in yours wmv videos by means of mac farewell ablaze wmv en route to dvd:  
 GtxRWm 0  
 Progressive greetings!

In about a month students would have set their pens ablaze in The Torch  
 Publications'... <http://t.co/9FxiXQuJt> 0  
 Rene Ablaze & Jacinta - Secret 2k13 (Fallen Skies Edit) - Mar 30 2013  
<https://t.co/7MLMsUzV1Z> 0  
 @Navista7 Steve these fires out here are something else! California is a  
 tinderbox - and this clown was setting my 'hood ablaze @News24680 1  
 #NowPlaying: Rene Ablaze & Ian Buff - Magnitude <http://t.co/Av2JSjfFtc> #EDM  
 0  
 @nxwestmidlands huge fire at Wholesale markets ablaze <http://t.co/rwzbFVNxER>  
 1  
 @ablaze what time does your talk go until? I don't know if I can make it due to  
 work. 0  
 'I can't have kids cuz I got in a bicycle accident & split my testicles.  
 it's impossible for me to have kids' MICHAEL YOU ARE THE FATHER 0  
 Accident on I-24 W #NashvilleTraffic. Traffic moving 8m slower than usual.  
<https://t.co/0GHk693EgJ> 1  
 Accident center lane blocked in #SantaClara on US-101 NB before Great America  
 Pkwy #BayArea #Traffic <http://t.co/pml0hZuRWR> 1  
<http://t.co/GKYe6gjTk5> Had a #personalinjury accident this summer? Read our  
 advice & see how a #solicitor can help #OtleyHour 1  
 #stlouis #caraccidentlawyer Speeding Among Top Causes of Teen Accidents  
<https://t.co/k4zoMOF319> <https://t.co/S2kXVM0cBA> Car Accident tee Ü\_ 1  
 Reported motor vehicle accident in Curry on Herman Rd near Stephenson involving  
 an overturned vehicle. Please use... <http://t.co/YbJezKuRW1> 1  
 BigRigRadio Live Accident Awareness 1  
 I-77 Mile Marker 31 South Mooresville Iredell Vehicle Accident Ramp Closed at  
 8/6 1:18 PM 1  
 RT @SleepJunkies: Sleeping pills double your risk of a car accident  
<http://t.co/7s9Nm1fiCT> 0  
 'By accident' they knew what was gon happen <https://t.co/Ysxun5vCeh> 0  
 Traffic accident N CABRILLO HWY/MAGELLAN AV MIR (08/06/15 11:03:58) 1  
 I-77 Mile Marker 31 to 40 South Mooresville Iredell Vehicle Accident Congestion  
 at 8/6 1:18 PM 1  
 the pastor was not in the scene of the accident...who was the owner of the  
 range rover ? 1  
 mom: 'we didn't get home as fast as we wished'  
 me: 'why is that?'

mom: 'there was an accident and some truck spilt mayonnaise all over ??????  
0  
I was in a horrible car accident this past Sunday. I'm finally able to get  
around. Thank you GOD?? 1  
Can wait to see how pissed Donnie is when I tell him I was in ANOTHER accident??  
0  
#TruckCrash Overturms On #FortWorth Interstate <http://t.co/Rs22LJ4qFp> Click here  
if you've been in a crash&gt;<http://t.co/Ld0unIYw4k> 1  
Accident in #Ashville on US 23 SB before SR 752 #traffic <http://t.co/hylMoOWgFI>  
1  
Carolina accident: Motorcyclist Dies in I-540 Crash With Car That Crossed  
Median: A motorcycle rider traveling... <http://t.co/p18lzRlmy6> 1  
FYI CAD:FYI: ;ACCIDENT PROPERTY DAMAGE;NHS;999 PINER RD/HORNDAL DR 1  
RT nAAYf: First accident in years. Turning onto Chandanee Magu from near MMA.  
Taxi rammed into me while I was halfway turned. Everyone conf Ū\_ 1  
Accident left lane blocked in #Manchester on Rt 293 NB before Eddy Rd stop and  
go traffic back to NH-3A delay of 4 mins #traffic 1  
;ACCIDENT PROPERTY DAMAGE; PINER RD/HORNDAL DR 1  
???? it was an accident <http://t.co/Oia5fxi4gM> 1  
FYI CAD:FYI: ;ACCIDENT PROPERTY DAMAGE;WPD;1600 S 17TH ST 1  
8/6/2015@2:09 PM: TRAFFIC ACCIDENT NO INJURY at 2781 WILLIS FOREMAN RD  
<http://t.co/VCKIT6EDEv> 1  
Aashiqui Actress Anu Aggarwal On Her Near-Fatal Accident <http://t.co/60tftp31LqW>  
1  
Suffield Alberta Accident <https://t.co/bPTmlF4P10> 1  
9 Mile backup on I-77 South...accident blocking the Right 2 Lanes at Exit 31  
Langtree Rd...consider NC 115 or NC 150 to NC 16 as alternate 1  
Has an accident changed your life? We will help you determine options that can  
financially support life care plans and on-going treatment. 0  
#BREAKING: there was a deadly motorcycle car accident that happened to  
#Hagerstown today. I'll have more details at 5 @Your4State. #WHAG 1  
@flowri were you marinading it or was it an accident? 0  
only had a car for not even a week and got in a fucking car accident .. Mfs  
can't fucking drive . 1

[ ]:

### 0.0.1 Optional: custom class weight

```
[ ]: train_labels = encoded_dataset_train['label']
print(np.bincount(train_labels))
class_weights = compute_class_weight(class_weight='balanced', classes=np.
unique(train_labels), y=list(train_labels))
print(class_weights)
```

[3907 2944]

[0.87675966 1.16355299]

```
[ ]: class CustomTrainer(Trainer):
    def compute_loss(self, model, inputs, return_outputs=False):
        #print(inputs)
        labels = inputs.get("labels")
        # forward pass
        outputs = model(**inputs)
        logits = outputs.get("logits")
        # compute custom loss (suppose one has 2 labels with different weights)
        loss_fct = nn.CrossEntropyLoss(weight=torch.Tensor(class_weights).
↳to('cuda'))
        loss = loss_fct(logits.view(-1, self.model.config.num_labels), labels.
↳view(-1))
        return (loss, outputs) if return_outputs else loss
```

```
[ ]: trainer = CustomTrainer(
    model,
    args,
    train_dataset=encoded_dataset_train,
    eval_dataset=encoded_dataset_dev,
    tokenizer=tokenizer,
    compute_metrics=compute_metrics
)
```

```
[ ]: trainer.train()
```

The following columns in the training set don't have a corresponding argument in `BertForSequenceClassification.forward` and have been ignored: text, \_\_index\_level\_0\_\_. If text, \_\_index\_level\_0\_\_ are not expected by `BertForSequenceClassification.forward`, you can safely ignore this message.  
/usr/local/lib/python3.7/dist-packages/transformers/optimization.py:309:  
FutureWarning: This implementation of AdamW is deprecated and will be removed in a future version. Use the PyTorch implementation torch.optim.AdamW instead, or set `no\_deprecation\_warning=True` to disable this warning

```
FutureWarning,
***** Running training *****
    Num examples = 6851
    Num Epochs = 4
    Instantaneous batch size per device = 8
    Total train batch size (w. parallel, distributed & accumulation) = 8
    Gradient Accumulation steps = 1
    Total optimization steps = 3428
```

<IPython.core.display.HTML object>

The following columns in the evaluation set don't have a corresponding argument in `BertForSequenceClassification.forward` and have been ignored: text, \_\_index\_level\_0\_\_. If text, \_\_index\_level\_0\_\_ are not expected by `BertForSequenceClassification.forward`, you can safely ignore this message.

```
***** Running Evaluation *****
```

```

    Num examples = 762
    Batch size = 8
Saving model checkpoint to ./save_model/bert-base-uncased-finetuned-
tweet/checkpoint-857
Configuration saved in ./save_model/bert-base-uncased-finetuned-
tweet/checkpoint-857/config.json
Model weights saved in ./save_model/bert-base-uncased-finetuned-
tweet/checkpoint-857/pytorch_model.bin
tokenizer config file saved in ./save_model/bert-base-uncased-finetuned-
tweet/checkpoint-857/tokenizer_config.json
Special tokens file saved in ./save_model/bert-base-uncased-finetuned-
tweet/checkpoint-857/special_tokens_map.json
The following columns in the evaluation set don't have a corresponding argument
in `BertForSequenceClassification.forward` and have been ignored: text,
__index_level_0__. If text, __index_level_0__ are not expected by
`BertForSequenceClassification.forward`, you can safely ignore this message.
***** Running Evaluation *****
    Num examples = 762
    Batch size = 8
Saving model checkpoint to ./save_model/bert-base-uncased-finetuned-
tweet/checkpoint-1714
Configuration saved in ./save_model/bert-base-uncased-finetuned-
tweet/checkpoint-1714/config.json
Model weights saved in ./save_model/bert-base-uncased-finetuned-
tweet/checkpoint-1714/pytorch_model.bin
tokenizer config file saved in ./save_model/bert-base-uncased-finetuned-
tweet/checkpoint-1714/tokenizer_config.json
Special tokens file saved in ./save_model/bert-base-uncased-finetuned-
tweet/checkpoint-1714/special_tokens_map.json
The following columns in the evaluation set don't have a corresponding argument
in `BertForSequenceClassification.forward` and have been ignored: text,
__index_level_0__. If text, __index_level_0__ are not expected by
`BertForSequenceClassification.forward`, you can safely ignore this message.
***** Running Evaluation *****
    Num examples = 762
    Batch size = 8
Saving model checkpoint to ./save_model/bert-base-uncased-finetuned-
tweet/checkpoint-2571
Configuration saved in ./save_model/bert-base-uncased-finetuned-
tweet/checkpoint-2571/config.json
Model weights saved in ./save_model/bert-base-uncased-finetuned-
tweet/checkpoint-2571/pytorch_model.bin
tokenizer config file saved in ./save_model/bert-base-uncased-finetuned-
tweet/checkpoint-2571/tokenizer_config.json
Special tokens file saved in ./save_model/bert-base-uncased-finetuned-
tweet/checkpoint-2571/special_tokens_map.json
The following columns in the evaluation set don't have a corresponding argument
in `BertForSequenceClassification.forward` and have been ignored: text,

```



\_\_index\_level\_0\_\_. If text, \_\_index\_level\_0\_\_ are not expected by  
`BertForSequenceClassification.forward`, you can safely ignore this message.

\*\*\*\*\* Running Evaluation \*\*\*\*\*

Num examples = 762

Batch size = 8

Saving model checkpoint to ./save\_model/bert-base-uncased-finetuned-  
tweet/checkpoint-3428

Configuration saved in ./save\_model/bert-base-uncased-finetuned-  
tweet/checkpoint-3428/config.json

Model weights saved in ./save\_model/bert-base-uncased-finetuned-  
tweet/checkpoint-3428/pytorch\_model.bin

tokenizer config file saved in ./save\_model/bert-base-uncased-finetuned-  
tweet/checkpoint-3428/tokenizer\_config.json

Special tokens file saved in ./save\_model/bert-base-uncased-finetuned-  
tweet/checkpoint-3428/special\_tokens\_map.json

Training completed. Do not forget to share your model on [huggingface.co/models](https://huggingface.co/models)  
=)

Loading best model from ./save\_model/bert-base-uncased-finetuned-  
tweet/checkpoint-2571 (score: 0.8267629388511841).

```
[ ]: TrainOutput(global_step=3428, training_loss=0.2242263940835639,  
  metrics={'train_runtime': 435.6999, 'train_samples_per_second': 62.897,  
    'train_steps_per_second': 7.868, 'total_flos': 709032391428420.0, 'train_loss':  
    0.2242263940835639, 'epoch': 4.0})
```

```
[ ]:
```