

bert_electra_albert

December 5, 2021

```
[1]: ! python -m pip install tf-models-official
! python -m pip install tensorflow
! python -m pip install tensorflow-text
! python -m pip install spacy
! python -m spacy download en_core_web_sm
! python -m spacy validate
```

Collecting tf-models-official

Downloading tf_models_official-2.7.0-py2.py3-none-any.whl (1.8 MB)

|| 1.8 MB 13.3 MB/s

Collecting tensorflow-addons

Downloading tensorflow_addons-0.15.0-cp37-cp37m-manylinux_2_12_x86_64.manylinux2010_x86_64.whl (1.1 MB)

|| 1.1 MB 36.7 MB/s

Requirement already satisfied: six in /usr/local/lib/python3.7/dist-packages (from tf-models-official) (1.15.0)

Requirement already satisfied: Pillow in /usr/local/lib/python3.7/dist-packages (from tf-models-official) (7.1.2)

Requirement already satisfied: numpy>=1.15.4 in /usr/local/lib/python3.7/dist-packages (from tf-models-official) (1.19.5)

Requirement already satisfied: tensorflow-hub>=0.6.0 in /usr/local/lib/python3.7/dist-packages (from tf-models-official) (0.12.0)

Requirement already satisfied: matplotlib in /usr/local/lib/python3.7/dist-packages (from tf-models-official) (3.2.2)

Collecting tensorflow-text>=2.7.0

Downloading tensorflow_text-2.7.3-cp37-cp37m-manylinux2010_x86_64.whl (4.9 MB)

|| 4.9 MB 11.8 MB/s

Requirement already satisfied: pandas>=0.22.0 in

/usr/local/lib/python3.7/dist-packages (from tf-models-official) (1.1.5)

Requirement already satisfied: gin-config in /usr/local/lib/python3.7/dist-packages (from tf-models-official) (0.5.0)

Collecting sacrebleu

Downloading sacrebleu-2.0.0-py3-none-any.whl (90 kB)

|| 90 kB 9.4 MB/s

Collecting pyyaml>=5.1

Downloading PyYAML-6.0-cp37-cp37m-manylinux_2_5_x86_64.manylinux1_x86_64.manylinux2_12_x86_64.manylinux2010_x86_64.whl (596 kB)

|| 596 kB 45.6 MB/s

Requirement already satisfied: Cython in /usr/local/lib/python3.7/dist-packages (from tf-models-official) (0.29.24)

Requirement already satisfied: google-api-python-client>=1.6.7 in /usr/local/lib/python3.7/dist-packages (from tf-models-official) (1.12.8)

Collecting sentencepiece

 Downloading

sentencepiece-0.1.96-cp37-cp37m-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (1.2 MB)

 || 1.2 MB 41.3 MB/s

Collecting py-cpuinfo>=3.3.0

 Downloading py-cpuinfo-8.0.0.tar.gz (99 kB)

 || 99 kB 9.2 MB/s

Requirement already satisfied: kaggle>=1.3.9 in /usr/local/lib/python3.7/dist-packages (from tf-models-official) (1.5.12)

Requirement already satisfied: psutil>=5.4.3 in /usr/local/lib/python3.7/dist-packages (from tf-models-official) (5.4.8)

Collecting opencv-python-headless

 Downloading opencv_python_headless-4.5.4.60-cp37-cp37m-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (47.6 MB)

 || 47.6 MB 101 kB/s

Collecting tensorflow-model-optimization>=0.4.1

 Downloading tensorflow_model_optimization-0.7.0-py2.py3-none-any.whl (213 kB)

 || 213 kB 43.7 MB/s

Requirement already satisfied: oauth2client in /usr/local/lib/python3.7/dist-packages (from tf-models-official) (4.1.3)

Collecting seqeval

 Downloading seqeval-1.2.2.tar.gz (43 kB)

 || 43 kB 2.2 MB/s

Requirement already satisfied: scipy>=0.19.1 in /usr/local/lib/python3.7/dist-packages (from tf-models-official) (1.4.1)

Requirement already satisfied: pycocotools in /usr/local/lib/python3.7/dist-packages (from tf-models-official) (2.0.3)

Collecting tf-slim>=1.1.0

 Downloading tf_slim-1.1.0-py2.py3-none-any.whl (352 kB)

 || 352 kB 34.0 MB/s

Requirement already satisfied: tensorflow-datasets in /usr/local/lib/python3.7/dist-packages (from tf-models-official) (4.0.1)

Requirement already satisfied: tensorflow>=2.7.0 in /usr/local/lib/python3.7/dist-packages (from tf-models-official) (2.7.0)

Requirement already satisfied: google-auth>=1.16.0 in /usr/local/lib/python3.7/dist-packages (from google-api-python-client>=1.6.7->tf-models-official) (1.35.0)

Requirement already satisfied: httplib2<1dev,>=0.15.0 in /usr/local/lib/python3.7/dist-packages (from google-api-python-client>=1.6.7->tf-models-official) (0.17.4)

Requirement already satisfied: uritemplate<4dev,>=3.0.0 in /usr/local/lib/python3.7/dist-packages (from google-api-python-client>=1.6.7->tf-models-official) (3.0.1)

Requirement already satisfied: google-api-core<2dev,>=1.21.0 in /usr/local/lib/python3.7/dist-packages (from google-api-python-client>=1.6.7->tf-models-official) (1.26.3)

Requirement already satisfied: google-auth-httpplib2>=0.0.3 in /usr/local/lib/python3.7/dist-packages (from google-api-python-client>=1.6.7->tf-models-official) (0.0.4)

Requirement already satisfied: pytz in /usr/local/lib/python3.7/dist-packages (from google-api-core<2dev,>=1.21.0->google-api-python-client>=1.6.7->tf-models-official) (2018.9)

Requirement already satisfied: protobuf>=3.12.0 in /usr/local/lib/python3.7/dist-packages (from google-api-core<2dev,>=1.21.0->google-api-python-client>=1.6.7->tf-models-official) (3.17.3)

Requirement already satisfied: packaging>=14.3 in /usr/local/lib/python3.7/dist-packages (from google-api-core<2dev,>=1.21.0->google-api-python-client>=1.6.7->tf-models-official) (21.3)

Requirement already satisfied: googleapis-common-protos<2.0dev,>=1.6.0 in /usr/local/lib/python3.7/dist-packages (from google-api-core<2dev,>=1.21.0->google-api-python-client>=1.6.7->tf-models-official) (1.53.0)

Requirement already satisfied: requests<3.0.0dev,>=2.18.0 in /usr/local/lib/python3.7/dist-packages (from google-api-core<2dev,>=1.21.0->google-api-python-client>=1.6.7->tf-models-official) (2.23.0)

Requirement already satisfied: setuptools>=40.3.0 in /usr/local/lib/python3.7/dist-packages (from google-api-core<2dev,>=1.21.0->google-api-python-client>=1.6.7->tf-models-official) (57.4.0)

Requirement already satisfied: pyasn1-modules>=0.2.1 in /usr/local/lib/python3.7/dist-packages (from google-auth>=1.16.0->google-api-python-client>=1.6.7->tf-models-official) (0.2.8)

Requirement already satisfied: cachetools<5.0,>=2.0.0 in /usr/local/lib/python3.7/dist-packages (from google-auth>=1.16.0->google-api-python-client>=1.6.7->tf-models-official) (4.2.4)

Requirement already satisfied: rsa<5,>=3.1.4 in /usr/local/lib/python3.7/dist-packages (from google-auth>=1.16.0->google-api-python-client>=1.6.7->tf-models-official) (4.8)

Requirement already satisfied: tqdm in /usr/local/lib/python3.7/dist-packages (from kaggle>=1.3.9->tf-models-official) (4.62.3)

Requirement already satisfied: certifi in /usr/local/lib/python3.7/dist-packages (from kaggle>=1.3.9->tf-models-official) (2021.10.8)

Requirement already satisfied: python-slugify in /usr/local/lib/python3.7/dist-packages (from kaggle>=1.3.9->tf-models-official) (5.0.2)

Requirement already satisfied: python-dateutil in /usr/local/lib/python3.7/dist-packages (from kaggle>=1.3.9->tf-models-official) (2.8.2)

Requirement already satisfied: urllib3 in /usr/local/lib/python3.7/dist-packages (from kaggle>=1.3.9->tf-models-official) (1.24.3)

Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in

/usr/local/lib/python3.7/dist-packages (from packaging>=14.3->google-api-core<2dev,>=1.21.0->google-api-python-client>=1.6.7->tf-models-official) (3.0.6)
 Requirement already satisfied: pyasn1<0.5.0,>=0.4.6 in
 /usr/local/lib/python3.7/dist-packages (from pyasn1-modules>=0.2.1->google-auth>=1.16.0->google-api-python-client>=1.6.7->tf-models-official) (0.4.8)
 Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-packages (from requests<3.0.0dev,>=2.18.0->google-api-core<2dev,>=1.21.0->google-api-python-client>=1.6.7->tf-models-official) (2.10)
 Requirement already satisfied: chardet<4,>=3.0.2 in
 /usr/local/lib/python3.7/dist-packages (from requests<3.0.0dev,>=2.18.0->google-api-core<2dev,>=1.21.0->google-api-python-client>=1.6.7->tf-models-official) (3.0.4)
 Requirement already satisfied: flatbuffers<3.0,>=1.12 in
 /usr/local/lib/python3.7/dist-packages (from tensorflow>=2.7.0->tf-models-official) (2.0)
 Requirement already satisfied: wrapt>=1.11.0 in /usr/local/lib/python3.7/dist-packages (from tensorflow>=2.7.0->tf-models-official) (1.13.3)
 Requirement already satisfied: absl-py>=0.4.0 in /usr/local/lib/python3.7/dist-packages (from tensorflow>=2.7.0->tf-models-official) (0.12.0)
 Requirement already satisfied: keras<2.8,>=2.7.0rc0 in
 /usr/local/lib/python3.7/dist-packages (from tensorflow>=2.7.0->tf-models-official) (2.7.0)
 Requirement already satisfied: gast<0.5.0,>=0.2.1 in
 /usr/local/lib/python3.7/dist-packages (from tensorflow>=2.7.0->tf-models-official) (0.4.0)
 Requirement already satisfied: tensorboard~=2.6 in
 /usr/local/lib/python3.7/dist-packages (from tensorflow>=2.7.0->tf-models-official) (2.7.0)
 Requirement already satisfied: astunparse>=1.6.0 in
 /usr/local/lib/python3.7/dist-packages (from tensorflow>=2.7.0->tf-models-official) (1.6.3)
 Requirement already satisfied: wheel<1.0,>=0.32.0 in
 /usr/local/lib/python3.7/dist-packages (from tensorflow>=2.7.0->tf-models-official) (0.37.0)
 Requirement already satisfied: h5py>=2.9.0 in /usr/local/lib/python3.7/dist-packages (from tensorflow>=2.7.0->tf-models-official) (3.1.0)
 Requirement already satisfied: tensorflow-io-gcs-filesystem>=0.21.0 in
 /usr/local/lib/python3.7/dist-packages (from tensorflow>=2.7.0->tf-models-official) (0.22.0)
 Requirement already satisfied: keras-preprocessing>=1.1.1 in
 /usr/local/lib/python3.7/dist-packages (from tensorflow>=2.7.0->tf-models-official) (1.1.2)
 Requirement already satisfied: termcolor>=1.1.0 in
 /usr/local/lib/python3.7/dist-packages (from tensorflow>=2.7.0->tf-models-official) (1.1.0)
 Requirement already satisfied: google-pasta>=0.1.1 in
 /usr/local/lib/python3.7/dist-packages (from tensorflow>=2.7.0->tf-models-official) (0.2.0)

Requirement already satisfied: typing-extensions>=3.6.6 in /usr/local/lib/python3.7/dist-packages (from tensorflow>=2.7.0->tf-models-official) (3.10.0.2)

Requirement already satisfied: tensorflow-estimator<2.8,~=2.7.0rc0 in /usr/local/lib/python3.7/dist-packages (from tensorflow>=2.7.0->tf-models-official) (2.7.0)

Requirement already satisfied: grpcio<2.0,>=1.24.3 in /usr/local/lib/python3.7/dist-packages (from tensorflow>=2.7.0->tf-models-official) (1.42.0)

Requirement already satisfied: libclang>=9.0.1 in /usr/local/lib/python3.7/dist-packages (from tensorflow>=2.7.0->tf-models-official) (12.0.0)

Requirement already satisfied: opt-einsum>=2.3.2 in /usr/local/lib/python3.7/dist-packages (from tensorflow>=2.7.0->tf-models-official) (3.3.0)

Requirement already satisfied: cached-property in /usr/local/lib/python3.7/dist-packages (from h5py>=2.9.0->tensorflow>=2.7.0->tf-models-official) (1.5.2)

Requirement already satisfied: markdown>=2.6.8 in /usr/local/lib/python3.7/dist-packages (from tensorboard~=2.6->tensorflow>=2.7.0->tf-models-official) (3.3.6)

Requirement already satisfied: tensorboard-data-server<0.7.0,>=0.6.0 in /usr/local/lib/python3.7/dist-packages (from tensorboard~=2.6->tensorflow>=2.7.0->tf-models-official) (0.6.1)

Requirement already satisfied: google-auth-oauthlib<0.5,>=0.4.1 in /usr/local/lib/python3.7/dist-packages (from tensorboard~=2.6->tensorflow>=2.7.0->tf-models-official) (0.4.6)

Requirement already satisfied: werkzeug>=0.11.15 in /usr/local/lib/python3.7/dist-packages (from tensorboard~=2.6->tensorflow>=2.7.0->tf-models-official) (1.0.1)

Requirement already satisfied: tensorboard-plugin-wit>=1.6.0 in /usr/local/lib/python3.7/dist-packages (from tensorboard~=2.6->tensorflow>=2.7.0->tf-models-official) (1.8.0)

Requirement already satisfied: requests-oauthlib>=0.7.0 in /usr/local/lib/python3.7/dist-packages (from google-auth-oauthlib<0.5,>=0.4.1->tensorboard~=2.6->tensorflow>=2.7.0->tf-models-official) (1.3.0)

Requirement already satisfied: importlib-metadata>=4.4 in /usr/local/lib/python3.7/dist-packages (from markdown>=2.6.8->tensorboard~=2.6->tensorflow>=2.7.0->tf-models-official) (4.8.2)

Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.7/dist-packages (from importlib-metadata>=4.4->markdown>=2.6.8->tensorboard~=2.6->tensorflow>=2.7.0->tf-models-official) (3.6.0)

Requirement already satisfied: oauthlib>=3.0.0 in /usr/local/lib/python3.7/dist-packages (from requests-oauthlib>=0.7.0->google-auth-oauthlib<0.5,>=0.4.1->tensorboard~=2.6->tensorflow>=2.7.0->tf-models-official) (3.1.1)

Requirement already satisfied: dm-tree~=0.1.1 in /usr/local/lib/python3.7/dist-packages (from tensorflow-model-optimization>=0.4.1->tf-models-official) (0.1.6)

Requirement already satisfied: cycycler>=0.10 in /usr/local/lib/python3.7/dist-packages (from matplotlib->tf-models-official) (0.11.0)

Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.7/dist-packages (from matplotlib->tf-models-official) (1.3.2)

Requirement already satisfied: text-unidecode>=1.3 in /usr/local/lib/python3.7/dist-packages (from python-slugify->kaggle>=1.3.9->tf-models-official) (1.3)

Requirement already satisfied: regex in /usr/local/lib/python3.7/dist-packages (from sacrebleu->tf-models-official) (2019.12.20)

Collecting colorama

 Downloading colorama-0.4.4-py2.py3-none-any.whl (16 kB)

Collecting portalocker

 Downloading portalocker-2.3.2-py2.py3-none-any.whl (15 kB)

Requirement already satisfied: tabulate>=0.8.9 in /usr/local/lib/python3.7/dist-packages (from sacrebleu->tf-models-official) (0.8.9)

Requirement already satisfied: scikit-learn>=0.21.3 in /usr/local/lib/python3.7/dist-packages (from sequeval->tf-models-official) (1.0.1)

Requirement already satisfied: joblib>=0.11 in /usr/local/lib/python3.7/dist-packages (from scikit-learn>=0.21.3->sequeval->tf-models-official) (1.1.0)

Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.7/dist-packages (from scikit-learn>=0.21.3->sequeval->tf-models-official) (3.0.0)

Requirement already satisfied: typeguard>=2.7 in /usr/local/lib/python3.7/dist-packages (from tensorflow-addons->tf-models-official) (2.7.1)

Requirement already satisfied: future in /usr/local/lib/python3.7/dist-packages (from tensorflow-datasets->tf-models-official) (0.16.0)

Requirement already satisfied: promise in /usr/local/lib/python3.7/dist-packages (from tensorflow-datasets->tf-models-official) (2.3)

Requirement already satisfied: importlib-resources in /usr/local/lib/python3.7/dist-packages (from tensorflow-datasets->tf-models-official) (5.4.0)

Requirement already satisfied: attrs>=18.1.0 in /usr/local/lib/python3.7/dist-packages (from tensorflow-datasets->tf-models-official) (21.2.0)

Requirement already satisfied: dill in /usr/local/lib/python3.7/dist-packages (from tensorflow-datasets->tf-models-official) (0.3.4)

Requirement already satisfied: tensorflow-metadata in /usr/local/lib/python3.7/dist-packages (from tensorflow-datasets->tf-models-official) (1.4.0)

Building wheels for collected packages: py-cpuinfo, sequeval

 Building wheel for py-cpuinfo (setup.py) ... done

 Created wheel for py-cpuinfo: filename=py_cpuinfo-8.0.0-py3-none-any.whl size=22258 sha256=67b616752c11468c376c8450c49345a17fa40415ff41ea0fd4a100c8672fa8dd

 Stored in directory: /root/.cache/pip/wheels/d2/f1/1f/041add21dc9c4220157f1bd2bd6afe1f1a49524c3396b94401

 Building wheel for sequeval (setup.py) ... done

Created wheel for sequeval: filename=sequeval-1.2.2-py3-none-any.whl size=16181 sha256=e588cf97d9a366106fc764eb4f5d5561c1a1177dd70b30b49dd314e72cfb6fed

Stored in directory: /root/.cache/pip/wheels/05/96/ee/7cac4e74f3b19e3158dce26a20a1c86b3533c43ec72a549fd7

Successfully built py-cpuinfo sequeval

Installing collected packages: portlocker, colorama, tf-slim, tensorflow-text, tensorflow-model-optimization, tensorflow-addons, sequeval, sentencepiece, sacrebleu, pyyaml, py-cpuinfo, opencv-python-headless, tf-models-official

Attempting uninstall: pyyaml

Found existing installation: PyYAML 3.13

Uninstalling PyYAML-3.13:

Successfully uninstalled PyYAML-3.13

Successfully installed colorama-0.4.4 opencv-python-headless-4.5.4.60 portlocker-2.3.2 py-cpuinfo-8.0.0 pyyaml-6.0 sacrebleu-2.0.0 sentencepiece-0.1.96 sequeval-1.2.2 tensorflow-addons-0.15.0 tensorflow-model-optimization-0.7.0 tensorflow-text-2.7.3 tf-models-official-2.7.0 tf-slim-1.1.0

Requirement already satisfied: tensorflow in /usr/local/lib/python3.7/dist-packages (2.7.0)

Requirement already satisfied: wheel<1.0,>=0.32.0 in /usr/local/lib/python3.7/dist-packages (from tensorflow) (0.37.0)

Requirement already satisfied: tensorflow-estimator<2.8,~=2.7.0rc0 in /usr/local/lib/python3.7/dist-packages (from tensorflow) (2.7.0)

Requirement already satisfied: wrapt>=1.11.0 in /usr/local/lib/python3.7/dist-packages (from tensorflow) (1.13.3)

Requirement already satisfied: h5py>=2.9.0 in /usr/local/lib/python3.7/dist-packages (from tensorflow) (3.1.0)

Requirement already satisfied: keras-preprocessing>=1.1.1 in /usr/local/lib/python3.7/dist-packages (from tensorflow) (1.1.2)

Requirement already satisfied: tensorboard~=2.6 in /usr/local/lib/python3.7/dist-packages (from tensorflow) (2.7.0)

Requirement already satisfied: protobuf>=3.9.2 in /usr/local/lib/python3.7/dist-packages (from tensorflow) (3.17.3)

Requirement already satisfied: termcolor>=1.1.0 in /usr/local/lib/python3.7/dist-packages (from tensorflow) (1.1.0)

Requirement already satisfied: opt-einsum>=2.3.2 in /usr/local/lib/python3.7/dist-packages (from tensorflow) (3.3.0)

Requirement already satisfied: flatbuffers<3.0,>=1.12 in /usr/local/lib/python3.7/dist-packages (from tensorflow) (2.0)

Requirement already satisfied: six>=1.12.0 in /usr/local/lib/python3.7/dist-packages (from tensorflow) (1.15.0)

Requirement already satisfied: typing-extensions>=3.6.6 in /usr/local/lib/python3.7/dist-packages (from tensorflow) (3.10.0.2)

Requirement already satisfied: numpy>=1.14.5 in /usr/local/lib/python3.7/dist-packages (from tensorflow) (1.19.5)

Requirement already satisfied: keras<2.8,>=2.7.0rc0 in /usr/local/lib/python3.7/dist-packages (from tensorflow) (2.7.0)

Requirement already satisfied: gast<0.5.0,>=0.2.1 in /usr/local/lib/python3.7/dist-packages (from tensorflow) (0.4.0)

Requirement already satisfied: tensorflow-io-gcs-filesystem>=0.21.0 in /usr/local/lib/python3.7/dist-packages (from tensorflow) (0.22.0)

Requirement already satisfied: grpcio<2.0,>=1.24.3 in /usr/local/lib/python3.7/dist-packages (from tensorflow) (1.42.0)

Requirement already satisfied: absl-py>=0.4.0 in /usr/local/lib/python3.7/dist-packages (from tensorflow) (0.12.0)

Requirement already satisfied: libclang>=9.0.1 in /usr/local/lib/python3.7/dist-packages (from tensorflow) (12.0.0)

Requirement already satisfied: google-pasta>=0.1.1 in /usr/local/lib/python3.7/dist-packages (from tensorflow) (0.2.0)

Requirement already satisfied: astunparse>=1.6.0 in /usr/local/lib/python3.7/dist-packages (from tensorflow) (1.6.3)

Requirement already satisfied: cached-property in /usr/local/lib/python3.7/dist-packages (from h5py>=2.9.0->tensorflow) (1.5.2)

Requirement already satisfied: werkzeug>=0.11.15 in /usr/local/lib/python3.7/dist-packages (from tensorboard~=2.6->tensorflow) (1.0.1)

Requirement already satisfied: markdown>=2.6.8 in /usr/local/lib/python3.7/dist-packages (from tensorboard~=2.6->tensorflow) (3.3.6)

Requirement already satisfied: google-auth-oauthlib<0.5,>=0.4.1 in /usr/local/lib/python3.7/dist-packages (from tensorboard~=2.6->tensorflow) (0.4.6)

Requirement already satisfied: google-auth<3,>=1.6.3 in /usr/local/lib/python3.7/dist-packages (from tensorboard~=2.6->tensorflow) (1.35.0)

Requirement already satisfied: tensorboard-plugin-wit>=1.6.0 in /usr/local/lib/python3.7/dist-packages (from tensorboard~=2.6->tensorflow) (1.8.0)

Requirement already satisfied: tensorboard-data-server<0.7.0,>=0.6.0 in /usr/local/lib/python3.7/dist-packages (from tensorboard~=2.6->tensorflow) (0.6.1)

Requirement already satisfied: setuptools>=41.0.0 in /usr/local/lib/python3.7/dist-packages (from tensorboard~=2.6->tensorflow) (57.4.0)

Requirement already satisfied: requests<3,>=2.21.0 in /usr/local/lib/python3.7/dist-packages (from tensorboard~=2.6->tensorflow) (2.23.0)

Requirement already satisfied: pyasn1-modules>=0.2.1 in /usr/local/lib/python3.7/dist-packages (from google-auth<3,>=1.6.3->tensorboard~=2.6->tensorflow) (0.2.8)

Requirement already satisfied: rsa<5,>=3.1.4 in /usr/local/lib/python3.7/dist-packages (from google-auth<3,>=1.6.3->tensorboard~=2.6->tensorflow) (4.8)

Requirement already satisfied: cachetools<5.0,>=2.0.0 in /usr/local/lib/python3.7/dist-packages (from google-auth<3,>=1.6.3->tensorboard~=2.6->tensorflow) (4.2.4)

Requirement already satisfied: requests-oauthlib>=0.7.0 in /usr/local/lib/python3.7/dist-packages (from google-auth-oauthlib<0.5,>=0.4.1->tensorboard~=2.6->tensorflow) (1.3.0)

Requirement already satisfied: importlib-metadata>=4.4 in /usr/local/lib/python3.7/dist-packages (from markdown>=2.6.8->tensorboard~=2.6->tensorflow) (4.8.2)

Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.7/dist-packages (from importlib-metadata>=4.4->markdown>=2.6.8->tensorboard~=2.6->tensorflow) (3.6.0)

Requirement already satisfied: pyasn1<0.5.0,>=0.4.6 in /usr/local/lib/python3.7/dist-packages (from pyasn1-modules>=0.2.1->google-auth<3,>=1.6.3->tensorboard~=2.6->tensorflow) (0.4.8)

Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/dist-packages (from requests<3,>=2.21.0->tensorboard~=2.6->tensorflow) (3.0.4)

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<3,>=2.21.0->tensorboard~=2.6->tensorflow) (1.24.3)

Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-packages (from requests<3,>=2.21.0->tensorboard~=2.6->tensorflow) (2.10)

Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-packages (from requests<3,>=2.21.0->tensorboard~=2.6->tensorflow) (2021.10.8)

Requirement already satisfied: oauthlib>=3.0.0 in /usr/local/lib/python3.7/dist-packages (from requests-oauthlib>=0.7.0->google-auth-oauthlib<0.5,>=0.4.1->tensorboard~=2.6->tensorflow) (3.1.1)

Requirement already satisfied: tensorflow-text in /usr/local/lib/python3.7/dist-packages (2.7.3)

Requirement already satisfied: tensorflow-hub>=0.8.0 in /usr/local/lib/python3.7/dist-packages (from tensorflow-text) (0.12.0)

Requirement already satisfied: tensorflow<2.8,>=2.7.0 in /usr/local/lib/python3.7/dist-packages (from tensorflow-text) (2.7.0)

Requirement already satisfied: protobuf>=3.9.2 in /usr/local/lib/python3.7/dist-packages (from tensorflow<2.8,>=2.7.0->tensorflow-text) (3.17.3)

Requirement already satisfied: numpy>=1.14.5 in /usr/local/lib/python3.7/dist-packages (from tensorflow<2.8,>=2.7.0->tensorflow-text) (1.19.5)

Requirement already satisfied: tensorflow-estimator<2.8,~=2.7.0rc0 in /usr/local/lib/python3.7/dist-packages (from tensorflow<2.8,>=2.7.0->tensorflow-text) (2.7.0)

Requirement already satisfied: six>=1.12.0 in /usr/local/lib/python3.7/dist-packages (from tensorflow<2.8,>=2.7.0->tensorflow-text) (1.15.0)

Requirement already satisfied: grpcio<2.0,>=1.24.3 in /usr/local/lib/python3.7/dist-packages (from tensorflow<2.8,>=2.7.0->tensorflow-text) (1.42.0)

Requirement already satisfied: wheel<1.0,>=0.32.0 in /usr/local/lib/python3.7/dist-packages (from tensorflow<2.8,>=2.7.0->tensorflow-text) (0.37.0)

Requirement already satisfied: libclang>=9.0.1 in /usr/local/lib/python3.7/dist-packages (from tensorflow<2.8,>=2.7.0->tensorflow-text) (12.0.0)

Requirement already satisfied: astunparse>=1.6.0 in /usr/local/lib/python3.7/dist-packages (from tensorflow<2.8,>=2.7.0->tensorflow-

text) (1.6.3)

Requirement already satisfied: typing-extensions>=3.6.6 in /usr/local/lib/python3.7/dist-packages (from tensorflow<2.8,>=2.7.0->tensorflow-text) (3.10.0.2)

Requirement already satisfied: wrapt>=1.11.0 in /usr/local/lib/python3.7/dist-packages (from tensorflow<2.8,>=2.7.0->tensorflow-text) (1.13.3)

Requirement already satisfied: h5py>=2.9.0 in /usr/local/lib/python3.7/dist-packages (from tensorflow<2.8,>=2.7.0->tensorflow-text) (3.1.0)

Requirement already satisfied: tensorflow-io-gcs-filesystem>=0.21.0 in /usr/local/lib/python3.7/dist-packages (from tensorflow<2.8,>=2.7.0->tensorflow-text) (0.22.0)

Requirement already satisfied: gast<0.5.0,>=0.2.1 in /usr/local/lib/python3.7/dist-packages (from tensorflow<2.8,>=2.7.0->tensorflow-text) (0.4.0)

Requirement already satisfied: absl-py>=0.4.0 in /usr/local/lib/python3.7/dist-packages (from tensorflow<2.8,>=2.7.0->tensorflow-text) (0.12.0)

Requirement already satisfied: google-pasta>=0.1.1 in /usr/local/lib/python3.7/dist-packages (from tensorflow<2.8,>=2.7.0->tensorflow-text) (0.2.0)

Requirement already satisfied: tensorboard~=2.6 in /usr/local/lib/python3.7/dist-packages (from tensorflow<2.8,>=2.7.0->tensorflow-text) (2.7.0)

Requirement already satisfied: flatbuffers<3.0,>=1.12 in /usr/local/lib/python3.7/dist-packages (from tensorflow<2.8,>=2.7.0->tensorflow-text) (2.0)

Requirement already satisfied: termcolor>=1.1.0 in /usr/local/lib/python3.7/dist-packages (from tensorflow<2.8,>=2.7.0->tensorflow-text) (1.1.0)

Requirement already satisfied: keras<2.8,>=2.7.0rc0 in /usr/local/lib/python3.7/dist-packages (from tensorflow<2.8,>=2.7.0->tensorflow-text) (2.7.0)

Requirement already satisfied: keras-preprocessing>=1.1.1 in /usr/local/lib/python3.7/dist-packages (from tensorflow<2.8,>=2.7.0->tensorflow-text) (1.1.2)

Requirement already satisfied: opt-einsum>=2.3.2 in /usr/local/lib/python3.7/dist-packages (from tensorflow<2.8,>=2.7.0->tensorflow-text) (3.3.0)

Requirement already satisfied: cached-property in /usr/local/lib/python3.7/dist-packages (from h5py>=2.9.0->tensorflow<2.8,>=2.7.0->tensorflow-text) (1.5.2)

Requirement already satisfied: requests<3,>=2.21.0 in /usr/local/lib/python3.7/dist-packages (from tensorboard~=2.6->tensorflow<2.8,>=2.7.0->tensorflow-text) (2.23.0)

Requirement already satisfied: google-auth-oauthlib<0.5,>=0.4.1 in /usr/local/lib/python3.7/dist-packages (from tensorboard~=2.6->tensorflow<2.8,>=2.7.0->tensorflow-text) (0.4.6)

Requirement already satisfied: setuptools>=41.0.0 in /usr/local/lib/python3.7/dist-packages (from tensorboard~=2.6->tensorflow<2.8,>=2.7.0->tensorflow-text) (57.4.0)

Requirement already satisfied: tensorboard-data-server<0.7.0,>=0.6.0 in /usr/local/lib/python3.7/dist-packages (from tensorboard~=2.6->tensorflow<2.8,>=2.7.0->tensorflow-text) (0.6.1)

Requirement already satisfied: tensorboard-plugin-wit>=1.6.0 in /usr/local/lib/python3.7/dist-packages (from tensorboard~=2.6->tensorflow<2.8,>=2.7.0->tensorflow-text) (1.8.0)

Requirement already satisfied: werkzeug>=0.11.15 in /usr/local/lib/python3.7/dist-packages (from tensorboard~=2.6->tensorflow<2.8,>=2.7.0->tensorflow-text) (1.0.1)

Requirement already satisfied: markdown>=2.6.8 in /usr/local/lib/python3.7/dist-packages (from tensorboard~=2.6->tensorflow<2.8,>=2.7.0->tensorflow-text) (3.3.6)

Requirement already satisfied: google-auth<3,>=1.6.3 in /usr/local/lib/python3.7/dist-packages (from tensorboard~=2.6->tensorflow<2.8,>=2.7.0->tensorflow-text) (1.35.0)

Requirement already satisfied: cachetools<5.0,>=2.0.0 in /usr/local/lib/python3.7/dist-packages (from google-auth<3,>=1.6.3->tensorboard~=2.6->tensorflow<2.8,>=2.7.0->tensorflow-text) (4.2.4)

Requirement already satisfied: pyasn1-modules>=0.2.1 in /usr/local/lib/python3.7/dist-packages (from google-auth<3,>=1.6.3->tensorboard~=2.6->tensorflow<2.8,>=2.7.0->tensorflow-text) (0.2.8)

Requirement already satisfied: rsa<5,>=3.1.4 in /usr/local/lib/python3.7/dist-packages (from google-auth<3,>=1.6.3->tensorboard~=2.6->tensorflow<2.8,>=2.7.0->tensorflow-text) (4.8)

Requirement already satisfied: requests-oauthlib>=0.7.0 in /usr/local/lib/python3.7/dist-packages (from google-auth-oauthlib<0.5,>=0.4.1->tensorboard~=2.6->tensorflow<2.8,>=2.7.0->tensorflow-text) (1.3.0)

Requirement already satisfied: importlib-metadata>=4.4 in /usr/local/lib/python3.7/dist-packages (from markdown>=2.6.8->tensorboard~=2.6->tensorflow<2.8,>=2.7.0->tensorflow-text) (4.8.2)

Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.7/dist-packages (from importlib-metadata>=4.4->markdown>=2.6.8->tensorboard~=2.6->tensorflow<2.8,>=2.7.0->tensorflow-text) (3.6.0)

Requirement already satisfied: pyasn1<0.5.0,>=0.4.6 in /usr/local/lib/python3.7/dist-packages (from pyasn1-modules>=0.2.1->google-auth<3,>=1.6.3->tensorboard~=2.6->tensorflow<2.8,>=2.7.0->tensorflow-text) (0.4.8)

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<3,>=2.21.0->tensorboard~=2.6->tensorflow<2.8,>=2.7.0->tensorflow-text) (1.24.3)

Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-packages (from requests<3,>=2.21.0->tensorboard~=2.6->tensorflow<2.8,>=2.7.0->tensorflow-text)

(2.10)
Requirement already satisfied: chardet<4,>=3.0.2 in
/usr/local/lib/python3.7/dist-packages (from
requests<3,>=2.21.0->tensorboard~=2.6->tensorflow<2.8,>=2.7.0->tensorflow-text)
(3.0.4)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.7/dist-packages (from
requests<3,>=2.21.0->tensorboard~=2.6->tensorflow<2.8,>=2.7.0->tensorflow-text)
(2021.10.8)
Requirement already satisfied: oauthlib>=3.0.0 in /usr/local/lib/python3.7/dist-
packages (from requests-oauthlib>=0.7.0->google-auth-
oauthlib<0.5,>=0.4.1->tensorboard~=2.6->tensorflow<2.8,>=2.7.0->tensorflow-text)
(3.1.1)
Requirement already satisfied: spacy in /usr/local/lib/python3.7/dist-packages
(2.2.4)
Requirement already satisfied: plac<1.2.0,>=0.9.6 in
/usr/local/lib/python3.7/dist-packages (from spacy) (1.1.3)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in
/usr/local/lib/python3.7/dist-packages (from spacy) (1.0.6)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in
/usr/local/lib/python3.7/dist-packages (from spacy) (4.62.3)
Requirement already satisfied: wasabi<1.1.0,>=0.4.0 in
/usr/local/lib/python3.7/dist-packages (from spacy) (0.8.2)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in
/usr/local/lib/python3.7/dist-packages (from spacy) (2.23.0)
Requirement already satisfied: catalogue<1.1.0,>=0.0.7 in
/usr/local/lib/python3.7/dist-packages (from spacy) (1.0.0)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in
/usr/local/lib/python3.7/dist-packages (from spacy) (3.0.6)
Requirement already satisfied: srsly<1.1.0,>=1.0.2 in
/usr/local/lib/python3.7/dist-packages (from spacy) (1.0.5)
Requirement already satisfied: setuptools in /usr/local/lib/python3.7/dist-
packages (from spacy) (57.4.0)
Requirement already satisfied: blis<0.5.0,>=0.4.0 in
/usr/local/lib/python3.7/dist-packages (from spacy) (0.4.1)
Requirement already satisfied: thinc==7.4.0 in /usr/local/lib/python3.7/dist-
packages (from spacy) (7.4.0)
Requirement already satisfied: numpy>=1.15.0 in /usr/local/lib/python3.7/dist-
packages (from spacy) (1.19.5)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in
/usr/local/lib/python3.7/dist-packages (from spacy) (2.0.6)
Requirement already satisfied: importlib-metadata>=0.20 in
/usr/local/lib/python3.7/dist-packages (from catalogue<1.1.0,>=0.0.7->spacy)
(4.8.2)
Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.7/dist-
packages (from importlib-metadata>=0.20->catalogue<1.1.0,>=0.0.7->spacy) (3.6.0)
Requirement already satisfied: typing-extensions>=3.6.4 in
/usr/local/lib/python3.7/dist-packages (from importlib-

```

metadata>=0.20->catalogue<1.1.0,>=0.0.7->spacy) (3.10.0.2)
Requirement already satisfied: chardet<4,>=3.0.2 in
/usr/local/lib/python3.7/dist-packages (from requests<3.0.0,>=2.13.0->spacy)
(3.0.4)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.7/dist-packages (from requests<3.0.0,>=2.13.0->spacy)
(2021.10.8)
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-
packages (from requests<3.0.0,>=2.13.0->spacy) (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<3.0.0,>=2.13.0->spacy)
(1.24.3)
Collecting en_core_web_sm==2.2.5
  Downloading https://github.com/explosion/spacy-
models/releases/download/en_core_web_sm-2.2.5/en_core_web_sm-2.2.5.tar.gz (12.0
MB)
    || 12.0 MB 13.2 MB/s
Requirement already satisfied: spacy>=2.2.2 in
/usr/local/lib/python3.7/dist-packages (from en_core_web_sm==2.2.5) (2.2.4)
Requirement already satisfied: wasabi<1.1.0,>=0.4.0 in
/usr/local/lib/python3.7/dist-packages (from
spacy>=2.2.2->en_core_web_sm==2.2.5) (0.8.2)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in
/usr/local/lib/python3.7/dist-packages (from
spacy>=2.2.2->en_core_web_sm==2.2.5) (2.23.0)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in
/usr/local/lib/python3.7/dist-packages (from
spacy>=2.2.2->en_core_web_sm==2.2.5) (2.0.6)
Requirement already satisfied: srsly<1.1.0,>=1.0.2 in
/usr/local/lib/python3.7/dist-packages (from
spacy>=2.2.2->en_core_web_sm==2.2.5) (1.0.5)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in
/usr/local/lib/python3.7/dist-packages (from
spacy>=2.2.2->en_core_web_sm==2.2.5) (4.62.3)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in
/usr/local/lib/python3.7/dist-packages (from
spacy>=2.2.2->en_core_web_sm==2.2.5) (3.0.6)
Requirement already satisfied: catalogue<1.1.0,>=0.0.7 in
/usr/local/lib/python3.7/dist-packages (from
spacy>=2.2.2->en_core_web_sm==2.2.5) (1.0.0)
Requirement already satisfied: thinc==7.4.0 in /usr/local/lib/python3.7/dist-
packages (from spacy>=2.2.2->en_core_web_sm==2.2.5) (7.4.0)
Requirement already satisfied: setuptools in /usr/local/lib/python3.7/dist-
packages (from spacy>=2.2.2->en_core_web_sm==2.2.5) (57.4.0)
Requirement already satisfied: numpy>=1.15.0 in /usr/local/lib/python3.7/dist-
packages (from spacy>=2.2.2->en_core_web_sm==2.2.5) (1.19.5)
Requirement already satisfied: plac<1.2.0,>=0.9.6 in
/usr/local/lib/python3.7/dist-packages (from

```

```

spacy>=2.2.2->en_core_web_sm==2.2.5) (1.1.3)
Requirement already satisfied: blis<0.5.0,>=0.4.0 in
/usr/local/lib/python3.7/dist-packages (from
spacy>=2.2.2->en_core_web_sm==2.2.5) (0.4.1)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in
/usr/local/lib/python3.7/dist-packages (from
spacy>=2.2.2->en_core_web_sm==2.2.5) (1.0.6)
Requirement already satisfied: importlib-metadata>=0.20 in
/usr/local/lib/python3.7/dist-packages (from
catalogue<1.1.0,>=0.0.7->spacy>=2.2.2->en_core_web_sm==2.2.5) (4.8.2)
Requirement already satisfied: typing-extensions>=3.6.4 in
/usr/local/lib/python3.7/dist-packages (from importlib-
metadata>=0.20->catalogue<1.1.0,>=0.0.7->spacy>=2.2.2->en_core_web_sm==2.2.5)
(3.10.0.2)
Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.7/dist-
packages (from importlib-
metadata>=0.20->catalogue<1.1.0,>=0.0.7->spacy>=2.2.2->en_core_web_sm==2.2.5)
(3.6.0)
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<3.0.0,>=2.13.0->spacy>=2.2.2->en_core_web_sm==2.2.5) (1.24.3)
Requirement already satisfied: chardet<4,>=3.0.2 in
/usr/local/lib/python3.7/dist-packages (from
requests<3.0.0,>=2.13.0->spacy>=2.2.2->en_core_web_sm==2.2.5) (3.0.4)
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-
packages (from requests<3.0.0,>=2.13.0->spacy>=2.2.2->en_core_web_sm==2.2.5)
(2.10)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.7/dist-packages (from
requests<3.0.0,>=2.13.0->spacy>=2.2.2->en_core_web_sm==2.2.5) (2021.10.8)
Download and installation successful
You can now load the model via spacy.load('en_core_web_sm')
Loaded compatibility table

```

```

===== Installed models (spaCy v2.2.4)
=====

```

```

spaCy installation: /usr/local/lib/python3.7/dist-packages/spacy

```

TYPE	NAME	MODEL	VERSION
package	en-core-web-sm	en_core_web_sm	2.2.5
link	en	en_core_web_sm	2.2.5

```

[2]: import numpy as np
import pandas as pd
import os

```

```
import time
from tqdm import tqdm
```

```
[3]: import spacy
from spacy.lang.en.stop_words import STOP_WORDS
from spacy.lang.en import English
from sklearn.model_selection import train_test_split
from sklearn.metrics.pairwise import cosine_similarity
import re
import string
from bs4 import BeautifulSoup as bs
# Model Training
import tensorflow as tf
import tensorflow_hub as hub
import tensorflow_text as text
from tensorflow.keras import layers, Model
from tensorflow.keras import regularizers
from tensorflow.keras.metrics import BinaryAccuracy
from tensorflow.keras.losses import BinaryCrossentropy
import official.nlp.optimization
from official.nlp.optimization import create_optimizer # AdamW optimizer
from sklearn.metrics import roc_curve, confusion_matrix, f1_score
# Visualization
import seaborn as sns
from wordcloud import WordCloud
import matplotlib.pyplot as plt
from matplotlib import rcParams
# Version
from platform import python_version
```

```
[4]: RANDOM_SEED = 77
nlp = spacy.load('en_core_web_sm')
pd.set_option('display.max_colwidth', None) # Expand DataFrame column width
rcParams['figure.figsize'] = (10, 6) # Custom plot dimensions
sns.set_theme(palette='muted', style='whitegrid') # Seaborn plot theme
```

```
[5]: try:
    from google.colab import drive
    drive.mount('/content/drive')
    in_colab = True
except:
    in_colab = False
```

Mounted at /content/drive

```
[6]: if in_colab:
    # input_path = '/content/drive/MyDrive/NLP Project/codes/input'
    input_path = '/content/drive/MyDrive/ColabNotebooks/disaster_tweets/input'
```

```

current_path = '/content/drive/MyDrive/ColabNotebooks/disaster_tweets'
path = input_path + '/train.csv'
path_test = input_path + '/test.csv'
else:
    path = '../CS6120Project/input/train.csv'
    path_test = '../CS6120Project/input/test.csv'

```

```

[7]: df = pd.read_csv(path)
      print(df.shape)
      df.head()

```

(7613, 5)

```

[7]:   id  ... target
0    1  ...      1
1    4  ...      1
2    5  ...      1
3    6  ...      1
4    7  ...      1

```

[5 rows x 5 columns]

```

[8]: df_test = pd.read_csv(path_test)
      print(df_test.shape)
      df_test.head()

```

(3263, 4)

```

[8]:   id  ...
text
0    0  ... Just
happened a terrible car crash
1    2  ... Heard about #earthquake is different
cities, stay safe everyone.
2    3  ... there is a forest fire at spot pond, geese are fleeing across the
street, I cannot save them all
3    9  ... Apocalypse
lighting. #Spokane #wildfires
4   11  ... Typhoon Soudelor
kills 28 in China and Taiwan

```

[5 rows x 4 columns]

1 Data Analysis

```
[9]: duplicates = df[df.duplicated(['text', 'target'], keep=False)]
print(f'Train Duplicate Entries (text, target): {len(duplicates)}')
duplicates.head()
```

Train Duplicate Entries (text, target): 157

```
[9]:      id  ... target
40    59  ...      0
48    68  ...      0
106  156  ...      0
115  165  ...      0
118  171  ...      0
```

[5 rows x 5 columns]

```
[10]: df.drop_duplicates(['text', 'target'], inplace=True, ignore_index=True)
print(df.shape, df_test.shape)
```

(7521, 5) (3263, 4)

```
[11]: new_duplicates = df[df.duplicated(['keyword', 'text'], keep=False)]
new_duplicates[['text', 'target']].sort_values(by='text')
print(f'Train Duplicate Entries (keyword, text): {len(new_duplicates)}')
df.drop(list(new_duplicates.index), inplace=True)
```

Train Duplicate Entries (keyword, text): 30

```
[12]: df = df.reset_index(drop=True)
df
```

```
[12]:      id  ... target
0       1  ...      1
1       4  ...      1
2       5  ...      1
3       6  ...      1
4       7  ...      1
...    ...  ...    ...
7486  10863  ...      1
7487  10864  ...      1
7488  10866  ...      1
7489  10869  ...      1
7490  10873  ...      1
```

[7491 rows x 5 columns]

```
[13]: df['target'].value_counts() / len(df)
```

```
[13]: 0    0.574022
      1    0.425978
      Name: target, dtype: float64
```

2 Preprocessing Text and EDA

```
[14]: text = df['text']
      target = df['target']
      test_text = df_test['text']

      # Print random samples from the training text
      for i in np.random.randint(500, size=5):
          print(f'Tweet #{i}: ', text[i], '=> Target: ', target[i])
```

```
Tweet #493: im feeling attacked http://t.co/91jvYCxXVi => Target: 0
Tweet #341: Well done Celtic Fingers crossed for Aberdeen tomorrow night!
Armageddon eh... ?? => Target: 0
Tweet #245: :StarMade: :Stardate 3: :Planetary Annihilation::
http://t.co/I2hHvIUtm via @YouTube => Target: 1
Tweet #183: Pakistan air ambulance helicopter crash kills nine
http://t.co/8E7rY8eBMf => Target: 1
Tweet #110: So i guess no one actually wants any free Aftershock TC... =>
Target: 0
```

```
[15]: lookup_dict = {
      'abt' : 'about',
      'afaik' : 'as far as i know',
      'bc' : 'because',
      'bfn' : 'bye for now',
      'bgd' : 'background',
      'bh' : 'blockhead',
      'br' : 'best regards',
      'btw' : 'by the way',
      'cc' : 'carbon copy',
      'chk' : 'check',
      'dam' : 'do not annoy me',
      'dd' : 'dear daughter',
      'df' : 'dear fiance',
      'ds' : 'dear son',
      'dyk' : 'did you know',
      'em' : 'email',
      'ema' : 'email address',
      'ftf' : 'face to face',
      'fb' : 'facebook',
      'ff' : 'follow friday',
      'fotd' : 'find of the day',
      'ftw' : 'for the win',
```

'fwiw' : 'for what it is worth',
'gts' : 'guess the song',
'hagn' : 'have a good night',
'hand' : 'have a nice day',
'hotd' : 'headline of the day',
'ht' : 'heard through',
'hth' : 'hope that helps',
'ic' : 'i see',
'icymi' : 'in case you missed it',
'idk' : 'i do not know',
'ig' : 'instagram',
'iirc' : 'if i remember correctly',
'imho' : 'in my humble opinion',
'imo' : 'in my opinion',
'irl' : 'in real life',
'iwsn' : 'i want sex now',
'jk' : 'just kidding',
'jsyk' : 'just so you know',
'jv' : 'joint venture',
'kk' : 'cool cool',
'kyso' : 'knock your socks off',
'lmao' : 'laugh my ass off',
'lmk' : 'let me know',
'lo' : 'little one',
'lol' : 'laugh out loud',
'mm' : 'music monday',
'mirl' : 'meet in real life',
'mrjn' : 'marijuana',
'nbd' : 'no big deal',
'nct' : 'nobody cares though',
'njoy' : 'enjoy',
'nsfw' : 'not safe for work',
'nts' : 'note to self',
'oh' : 'overheard',
'omg' : 'oh my god',
'oomf' : 'one of my friends',
'orly' : 'oh really',
'plmk' : 'please let me know',
'pnp' : 'party and play',
'qotd' : 'quote of the day',
're' : 'in reply to in regards to',
'rtq' : 'read the question',
'rt' : 'retweet',
'sfw' : 'safe for work',
'smdh' : 'shaking my damn head',
'smh' : 'shaking my head',
'so' : 'significant other',

}

→ \S+) ')

```

digit_pattern = re.compile(r'[\d]+')
# From https://gist.github.com/slowkow/7a7f61f495e3dbb7e3d767f97bd7304
emoji_pattern = re.compile("[
    u"\U0001F600-\U0001F64F" # emoticons
    u"\U0001F300-\U0001F5FF" # symbols & pictographs
    u"\U0001F680-\U0001F6FF" # transport & map
    u"\U0001F1E0-\U0001F1FF" # flags (iOS)
    u"\U00002500-\U00002BEF" # chinese char
    u"\U00002702-\U000027B0"
    u"\U00002702-\U000027B0"
    u"\U000024C2-\U0001F251"
    u"\U0001f926-\U0001f937"
    u"\U00010000-\U0010ffff"
    u"\u2640-\u2642"
    u"\u2600-\u2B55"
    u"\u200d"
    u"\u23cf"
    u"\u23e9"
    u"\u231a"
    u"\ufe0f" # dingbats
    u"\u3030"
"]+", flags=re.UNICODE)

# Remove urls
url_strip = text_data.apply(lambda x: re.sub(url_pattern, '', x) if pd.
→isna(x) != True else x)
# Parse the HTML
html_parse = url_strip.apply(lambda x: bs(x, 'html.parser').get_text() if pd.
→isna(x) != True else x)
# Remove retweets
retweet_strip = html_parse.apply(lambda x: re.sub(retweet_pattern, '', x) if
→pd.isna(x) != True else x)
# Remove emojis
emoji_strip = retweet_strip.apply(lambda x: re.sub(emoji_pattern, '', x) if
→pd.isna(x) != True else x)
# Remove entities
entity_strip = emoji_strip.apply(lambda x: re.sub(entity_pattern, '', x) if
→pd.isna(x) != True else x)
# Lowercase the strings
lowercase = entity_strip.apply(lambda x: str.lower(x) if pd.isna(x) != True
→else x)
# Remove punctuation
punct_strip = lowercase.apply(lambda x: re.sub(f'[{re.escape(string.
→punctuation)}]', '', x) if pd.isna(x) != True else x)
# Convert abbreviations to words

```

```

abbrev_converted = punct_strip.apply(lambda x: abbrev_conversion(x) if pd.
→isna(x) != True else x)
# Remove digits
digit_strip = abbrev_converted.apply(lambda x: re.sub(digit_pattern, '', x)
→if pd.isna(x) != True else x)
# Lemmatize text and filter stopwords
lemma_and_stop = digit_strip.apply(lambda x: lemmatize_text(x) if pd.isna(x)
→!= True else x)

return lemma_and_stop

```

```

[17]: clean_text = np.asarray(standardize_text(text))
test_clean_text = np.asarray(standardize_text(test_text))

# Print random samples from the cleaned training text
for i in np.random.randint(500, size=5):
    print(f'Tweet #{i}: ', clean_text[i], '=> Target: ', target[i])

```

```

Tweet #368:  beyonce pick fan army beyhive => Target:  0
Tweet #169:  air ambulance scene crash car lorry emsne => Target:  1
Tweet #167:  early wake sister beg come ride wher ambulance hospital rodkiaai =>
Target:  1
Tweet #68:  accident center lane block santaclara nb great america pkwy bayarea
traffic => Target:  1
Tweet #471:  fact atomic bomb call little boy fat man say lot mentality go
attack => Target:  1

```

3 Data Preparation for Training

```

[18]: df['clean_text'] = pd.DataFrame(clean_text)
df_test['clean_text'] = pd.DataFrame(test_clean_text)

```

```

[19]: df['tweet_len'] = df['clean_text'].apply(lambda x: len(x))

```

```

[20]: # Load the sentence encoder
sentence_enc = hub.load('https://tfhub.dev/google/universal-sentence-encoder/4')

```

```

[21]: def extract_keywords(text, nlp=nlp):
    potential_keywords = []
    TOP_KEYWORD = -1
    # Create a list for keyword parts of speech
    pos_tag = ['ADJ', 'NOUN', 'PROPN']
    doc = nlp(text)

    for i in doc:
        if i.pos_ in pos_tag:

```

```

        potential_keywords.append(i.text)

    document_embed = sentence_enc([text])
    potential_embed = sentence_enc(potential_keywords)

    vector_distances = cosine_similarity(document_embed, potential_embed)
    keyword = [potential_keywords[i] for i in vector_distances.
→argsort()[0][TOP_KEYWORD:]]

    return keyword

def keyword_filler(keyword, text):
    if pd.isnull(keyword):
        try:
            keyword = extract_keywords(text)[0]
        except:
            keyword = ''

    return keyword

```

```

[22]: df['keyword_fill'] = pd.DataFrame(list(map(keyword_filler, df['keyword'],
→df['clean_text']))).astype(str)
df_test['keyword_fill'] = pd.DataFrame(list(map(keyword_filler,
→df_test['keyword'], df_test['clean_text']))).astype(str)

print('Null Training Keywords => ', df['keyword_fill'].isnull().any())
print('Null Test Keywords => ', df_test['keyword_fill'].isnull().any())

```

```

Null Training Keywords => False
Null Test Keywords => False

```

```

[23]: df['keyword_fill'] = pd.DataFrame(standardize_text(df['keyword_fill']))
df_test['keyword_fill'] = pd.DataFrame(standardize_text(df_test['keyword_fill']))

```

```

[24]: df.head()

```

```

[24]:   id keyword  ... tweet_len keyword_fill
0    1   NaN  ...      36   earthquake
1    4   NaN  ...      37     forest
2    5   NaN  ...      79     shelter
3    6   NaN  ...      53    wildfire
4    7   NaN  ...      53    wildfire

[5 rows x 8 columns]

```

```

[25]: df_test.head()

```


[illegible]

```
train_features[:5]
```

```
clean_text keyword_fill
0 deed reason earthquake allah
forgive earthquake
1 forest fire near la ronge sask
canada forest
2 resident ask shelter place notify officer evacuation shelter place order
expect shelter
3 people receive wildfire evacuation order
california wildfire
4 get send photo ruby alaska smoke wildfire pour
school wildfire
```

```
test_features[:5]
```

	clean_text	keyword_fill
0	happen terrible car crash	crash
1	hear earthquake different city stay safe	earthquake
2	forest fire spot pond geese flee street save	geese
3	apocalypse light spokane wildfire	wildfire
4	typhoon soudelor kill china taiwan	taiwan

```
print(train_features.shape)
print(test_features.shape)
```

$$\begin{pmatrix} 7491 \\ 3263 \end{pmatrix}, 2$$

```
[30]: train_x, val_x, train_y, val_y = train_test_split(
        train_features,
        target,
        test_size=0.2,
        random_state=RANDOM_SEED,
    )

    print(train_x.shape)
    print(train_y.shape)
    print(val_x.shape)
    print(val_y.shape)
```

```
(5992, 2)
(5992,)
(1499, 2)
(1499,)
```

```
[31]: # Create TensorFlow Datasets
train_ds = tf.data.Dataset.from_tensor_slices((dict(train_x), train_y))
val_ds = tf.data.Dataset.from_tensor_slices((dict(val_x), val_y))
test_ds = tf.data.Dataset.from_tensor_slices(dict(test_features))
```

```
[32]: AUTOTUNE = tf.data.experimental.AUTOTUNE

BUFFER_SIZE = 1000
BATCH_SIZE = 32

def configure_dataset(dataset, shuffle=False, test=False):
    # Configure the tf dataset for cache, shuffle, batch, and prefetch
    if shuffle:
        dataset = dataset.cache()\
            .shuffle(BUFFER_SIZE, seed=RANDOM_SEED,
→reshuffle_each_iteration=True)\
            .batch(BATCH_SIZE, drop_remainder=True).
→prefetch(AUTOTUNE)
    elif test:
        dataset = dataset.cache()\
            .batch(BATCH_SIZE, drop_remainder=False).
→prefetch(AUTOTUNE)
    else:
        dataset = dataset.cache()\
            .batch(BATCH_SIZE, drop_remainder=True).
→prefetch(AUTOTUNE)
    return dataset
```

```
[33]: # Configure the datasets
train_ds = configure_dataset(train_ds, shuffle=True)
```

```
val_ds = configure_dataset(val_ds)
test_ds = configure_dataset(test_ds, test=True)
```

```
[34]: # Print the dataset specifications
```

```
print(train_ds.element_spec)
print(val_ds.element_spec)
print(test_ds.element_spec)
```

```
({'clean_text': TensorSpec(shape=(32,), dtype=tf.string, name=None),
'keyword_fill': TensorSpec(shape=(32,), dtype=tf.string, name=None)},
TensorSpec(shape=(32,), dtype=tf.int64, name=None))
({'clean_text': TensorSpec(shape=(32,), dtype=tf.string, name=None),
'keyword_fill': TensorSpec(shape=(32,), dtype=tf.string, name=None)},
TensorSpec(shape=(32,), dtype=tf.int64, name=None))
{'clean_text': TensorSpec(shape=(None,), dtype=tf.string, name=None),
'keyword_fill': TensorSpec(shape=(None,), dtype=tf.string, name=None)}
```

4 Building the Classifier Model

```
[35]: result = dict()
```

```
[36]: # BERT encoder w/ preprocessor
```

```
bert_preprocessor = hub.KerasLayer('https://tfhub.dev/tensorflow/
↳bert_en_uncased_preprocess/3', name='BERT_preprocessor')
# loading pre-trained BERT model with 12 hidden layers, 768 hidden unit size,
↳and 12 attention heads
bert_encoder = hub.KerasLayer('https://tfhub.dev/tensorflow/
↳bert_en_uncased_L-12_H-768_A-12/4', trainable=True, name='BERT_encoder')
# Keyword embedding layer
nnlm_embed = hub.KerasLayer('https://tfhub.dev/google/nnlm-en-dim50/2',
↳name='embedding_layer')
```

```
[37]: def build_model(preprocessor, encoder, name):
```

```
    # Construct text layers
    text_input = layers.Input(shape=(), dtype=tf.string, name='clean_text') #
    ↳Name matches df heading
    encoder_inputs = preprocessor(text_input)
    encoder_outputs = encoder(encoder_inputs)
    # pooled_output returns [batch_size, hidden_layers]
    pooled_output = encoder_outputs["pooled_output"]
    bert_dropout = layers.Dropout(0.1, name='BERT_dropout')(pooled_output)

    # Construct keyword layers
    key_input = layers.Input(shape=(), dtype=tf.string, name='keyword_fill') #
    ↳Name matches df heading
```

```

key_embed = nnlm_embed(key_input)
key_flat = layers.Flatten()(key_embed)
key_dense = layers.Dense(128, activation='elu',
→kernel_regularizer=regularizers.l2(1e-4))(key_flat)
key_dropout = layers.Dropout(0.5, name='dense_dropout')(key_dense)

# Merge the layers and classify
merge = layers.concatenate([bert_dropout, key_dropout])
dense = layers.Dense(128, activation='elu', kernel_regularizer=regularizers.
→l2(1e-4))(merge)
dropout = layers.Dropout(0.5, name='merged_dropout')(dense)
clf = layers.Dense(1, activation='sigmoid', name='classifier')(dropout)

return Model([text_input, key_input], clf, name=name)

```

```
[ ]: # tf.keras.utils.plot_model(bert_classifier, show_shapes=False, dpi=96)
```

4.0.1 BERT training

```
[ ]: # construction of AdamW optimizer
EPOCHS = 3
LEARNING_RATE = 5e-5

STEPS_PER_EPOCH = int(train_ds.unbatch().cardinality().numpy() / BATCH_SIZE)
VAL_STEPS = int(val_ds.unbatch().cardinality().numpy() / BATCH_SIZE)
# Calculate the train and warmup steps for the optimizer
TRAIN_STEPS = STEPS_PER_EPOCH * EPOCHS
WARMUP_STEPS = int(TRAIN_STEPS * 0.1)

adamw_optimizer = create_optimizer(
    init_lr=LEARNING_RATE,
    num_train_steps=TRAIN_STEPS,
    num_warmup_steps=WARMUP_STEPS
)

model_name = 'BERT'
bert_preprocessor = hub.KerasLayer('https://tfhub.dev/tensorflow/
→bert_en_uncased_preprocess/3', name='BERT_preprocessor')
# loading pre-trained BERT model with 12 hidden layers, 768 hidden unit size,
→and 12 attention heads
bert_encoder = hub.KerasLayer('https://tfhub.dev/tensorflow/
→bert_en_uncased_L-12_H-768_A-12/4', trainable=True, name='BERT_encoder')
bert_classifier = build_model(bert_preprocessor, bert_encoder, model_name)
bert_classifier.compile(
    loss=BinaryCrossentropy(from_logits=True),
    optimizer= adamw_optimizer,

```

```

        metrics=[BinaryAccuracy(name='accuracy')]
    )

    start_time = time.time()
    history_bert = bert_classifier.fit(
        train_ds,
        epochs=EPOCHS,
        steps_per_epoch=STEPS_PER_EPOCH,
        validation_data= val_ds,
        validation_steps=VAL_STEPS
    )
    end_time = time.time()

```

Epoch 1/3

```

/usr/local/lib/python3.7/dist-packages/tensorflow/python/util/dispatch.py:1096:
UserWarning: "`binary_crossentropy` received `from_logits=True`, but the
`output` argument was produced by a sigmoid or softmax activation and thus does
not represent logits. Was this intended?"

```

```

    return dispatch_target(*args, **kwargs)

```

```

187/187 [=====] - 365s 2s/step - loss: 0.5541 -
accuracy: 0.7520 - val_loss: 0.4633 - val_accuracy: 0.8077

```

Epoch 2/3

```

187/187 [=====] - 341s 2s/step - loss: 0.3857 -
accuracy: 0.8595 - val_loss: 0.4918 - val_accuracy: 0.8118

```

Epoch 3/3

```

187/187 [=====] - 340s 2s/step - loss: 0.2775 -
accuracy: 0.9084 - val_loss: 0.5438 - val_accuracy: 0.8091

```

```

[ ]: result[model_name] = {'training_time': end_time-start_time}
result[model_name]['val_accuracy'] = max(history_bert.history['val_accuracy'])
val_target = np.asarray([i[1] for i in list(val_ds.unbatch().
    ↳as_numpy_iterator())])
val_predict = bert_classifier.predict(val_ds)
val_predict = val_predict > 0.5
result[model_name]['F1_score'] = f1_score(val_target, val_predict)

```

4.0.2 Electra Training

```

[ ]: EPOCHS = 3
LEARNING_RATE = 1e-5

STEPS_PER_EPOCH = int(train_ds.unbatch().cardinality().numpy() / BATCH_SIZE)
VAL_STEPS = int(val_ds.unbatch().cardinality().numpy() / BATCH_SIZE)
# Calculate the train and warmup steps for the optimizer
TRAIN_STEPS = STEPS_PER_EPOCH * EPOCHS
WARMUP_STEPS = int(TRAIN_STEPS * 0.1)

```

```

adamw_optimizer = create_optimizer(
    init_lr=LEARNING_RATE,
    num_train_steps=TRAIN_STEPS,
    num_warmup_steps=WARMUP_STEPS
)

electra_preprocessor = hub.KerasLayer("https://tfhub.dev/tensorflow/
    ↪bert_en_uncased_preprocess/3")
electra_encoder = hub.KerasLayer("https://tfhub.dev/google/electra_small/2",
    ↪trainable=True, name="Electra_encoder")
model_name = 'Electra-small'
electra_classifier = build_model(electra_preprocessor, electra_encoder,
    ↪model_name)
electra_classifier.compile(
    loss=BinaryCrossentropy(from_logits=True),
    optimizer= adamw_optimizer,
    metrics=[BinaryAccuracy(name='accuracy')]
)

start_time = time.time()
history_electra = electra_classifier.fit(
    train_ds,
    epochs=EPOCHS,
    steps_per_epoch=STEPS_PER_EPOCH,
    validation_data= val_ds,
    validation_steps=VAL_STEPS
)
end_time = time.time()

```

Epoch 1/3

```

/usr/local/lib/python3.7/dist-packages/tensorflow/python/util/dispatch.py:1096:
UserWarning: ``binary_crossentropy` received `from_logits=True`, but the
`output` argument was produced by a sigmoid or softmax activation and thus does
not represent logits. Was this intended?"
    return dispatch_target(*args, **kwargs)

```

```

187/187 [=====] - 139s 630ms/step - loss: 0.7125 -
accuracy: 0.6093 - val_loss: 0.5643 - val_accuracy: 0.7303

```

Epoch 2/3

```

187/187 [=====] - 117s 625ms/step - loss: 0.5615 -
accuracy: 0.7572 - val_loss: 0.5407 - val_accuracy: 0.7656

```

Epoch 3/3

```

187/187 [=====] - 117s 625ms/step - loss: 0.5308 -
accuracy: 0.7821 - val_loss: 0.5519 - val_accuracy: 0.7507

```

```
[ ]: result[model_name] = {'training_time': end_time-start_time}
result[model_name]['val_accuracy'] = max(history_electra.history['val_accuracy'])
val_target = np.asarray([i[1] for i in list(val_ds.unbatch().
    ↳as_numpy_iterator())])
val_predict = electra_classifier.predict(val_ds)
val_predict = val_predict > 0.5
result[model_name]['F1_score'] = f1_score(val_target, val_predict)
```

4.0.3 ALBERT training

```
[38]: EPOCHS = 3
LEARNING_RATE = 1e-5

STEPS_PER_EPOCH = int(train_ds.unbatch().cardinality().numpy() / BATCH_SIZE)
VAL_STEPS = int(val_ds.unbatch().cardinality().numpy() / BATCH_SIZE)
# Calculate the train and warmup steps for the optimizer
TRAIN_STEPS = STEPS_PER_EPOCH * EPOCHS
WARMUP_STEPS = int(TRAIN_STEPS * 0.1)

adamw_optimizer = create_optimizer(
    init_lr=LEARNING_RATE,
    num_train_steps=TRAIN_STEPS,
    num_warmup_steps=WARMUP_STEPS
)

albert_preprocessor = hub.KerasLayer("http://tfhub.dev/tensorflow/
    ↳albert_en_preprocess/3")
albert_encoder = hub.KerasLayer("https://tfhub.dev/tensorflow/albert_en_base/3",
    ↳trainable=True, name="albert_encoder")
model_name = 'Albert-base'
albert_classifier = build_model(albert_preprocessor, albert_encoder, model_name)
albert_classifier.compile(
    loss=BinaryCrossentropy(from_logits=True),
    optimizer= adamw_optimizer,
    metrics=[BinaryAccuracy(name='accuracy')]
)

start_time = time.time()
history_albert = albert_classifier.fit(
    train_ds,
    epochs=EPOCHS,
    steps_per_epoch=STEPS_PER_EPOCH,
    validation_data= val_ds,
    validation_steps=VAL_STEPS
)
end_time = time.time()
```

Epoch 1/3

```
/usr/local/lib/python3.7/dist-packages/tensorflow/python/util/dispatch.py:1096:  
UserWarning: "`binary_crossentropy` received `from_logits=True`, but the  
`output` argument was produced by a sigmoid or softmax activation and thus does  
not represent logits. Was this intended?"
```

```
    return dispatch_target(*args, **kwargs)
```

```
187/187 [=====] - 391s 2s/step - loss: 0.7143 -  
accuracy: 0.6738 - val_loss: 0.5201 - val_accuracy: 0.7724
```

Epoch 2/3

```
187/187 [=====] - 377s 2s/step - loss: 0.5207 -  
accuracy: 0.7782 - val_loss: 0.5086 - val_accuracy: 0.7935
```

Epoch 3/3

```
187/187 [=====] - 377s 2s/step - loss: 0.4510 -  
accuracy: 0.8185 - val_loss: 0.5019 - val_accuracy: 0.7955
```

```
[39]: result[model_name] = {'training_time': end_time-start_time}  
result[model_name]['val_accuracy'] = max(history_albert.history['val_accuracy'])  
val_target = np.asarray([i[1] for i in list(val_ds.unbatch().  
    ↳as_numpy_iterator())])  
val_predict = albert_classifier.predict(val_ds)  
val_predict = val_predict > 0.5  
result[model_name]['F1_score'] = f1_score(val_target, val_predict)
```

```
[ ]: result
```

```
[ ]: {'Albert-base': {'F1_score': 0.7614125753660638,  
    'training_time': 1156.5787959098816,  
    'val_accuracy': 0.8118206262588501},  
    'BERT': {'F1_score': 0.7683429513602638,  
    'training_time': 1125.6142642498016,  
    'val_accuracy': 0.8118206262588501},  
    'Electra-small': {'F1_score': 0.7246811702925732,  
    'training_time': 422.6679632663727,  
    'val_accuracy': 0.765625}}
```

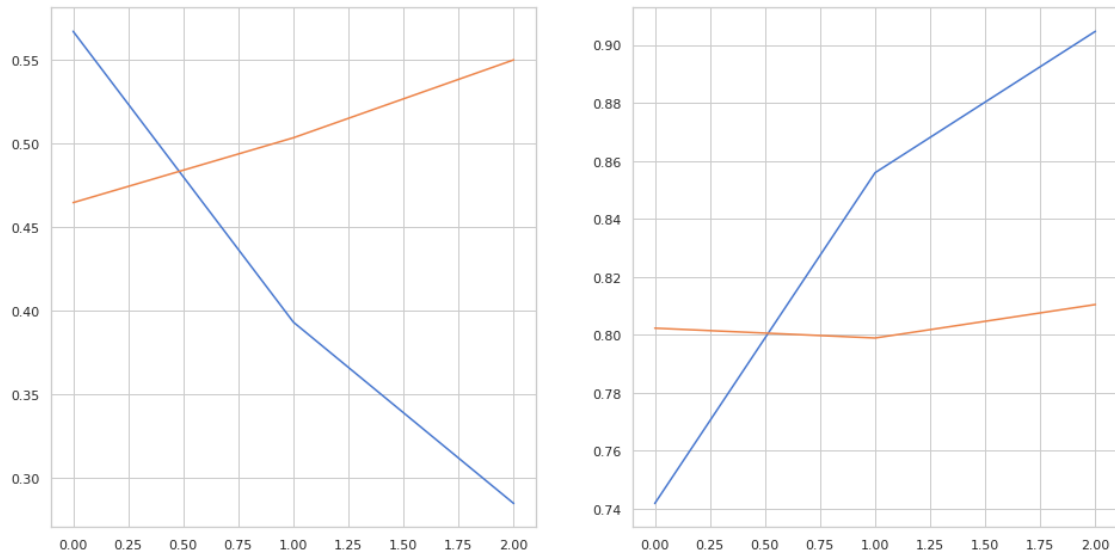
```
[ ]: df_result = pd.DataFrame(result)  
df_result = pd.DataFrame.transpose(df_result)  
df_result.to_csv(current_path + '/results.csv')
```

5 History plot, ROC curve plot and Confusion matrix for BERT classifier

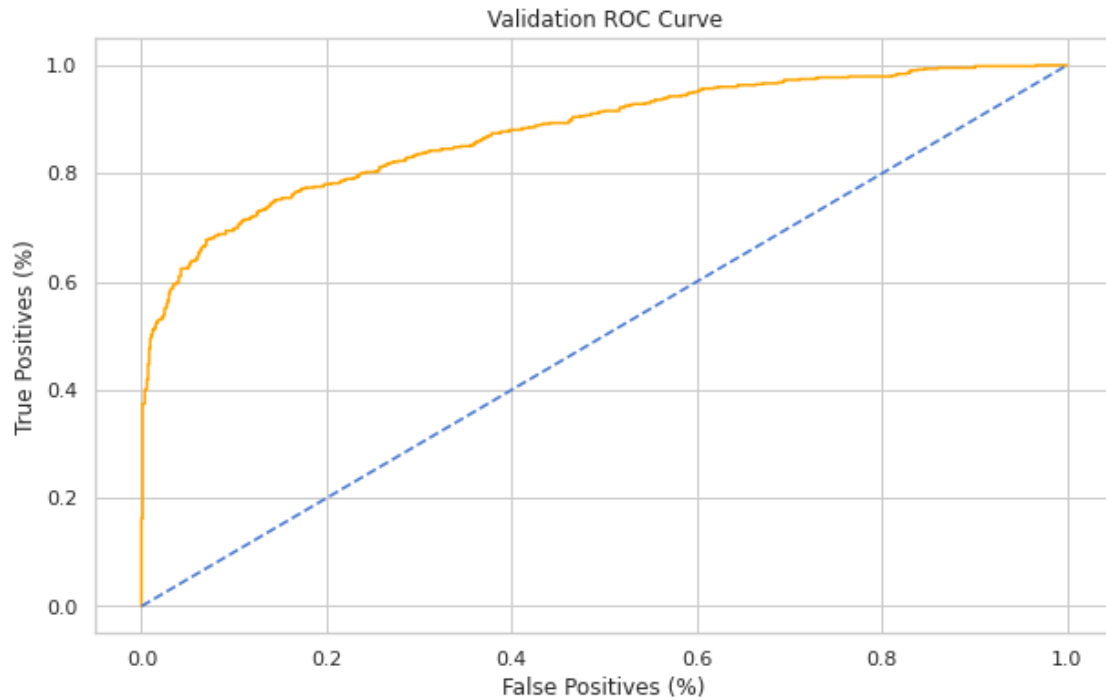
```
[ ]: def plot_history(history):  
    fig, ax = plt.subplots(1, 2, figsize=(16, 8))  
    ax[0].plot(history.history['loss'])  
    ax[0].plot(history.history['val_loss'])
```



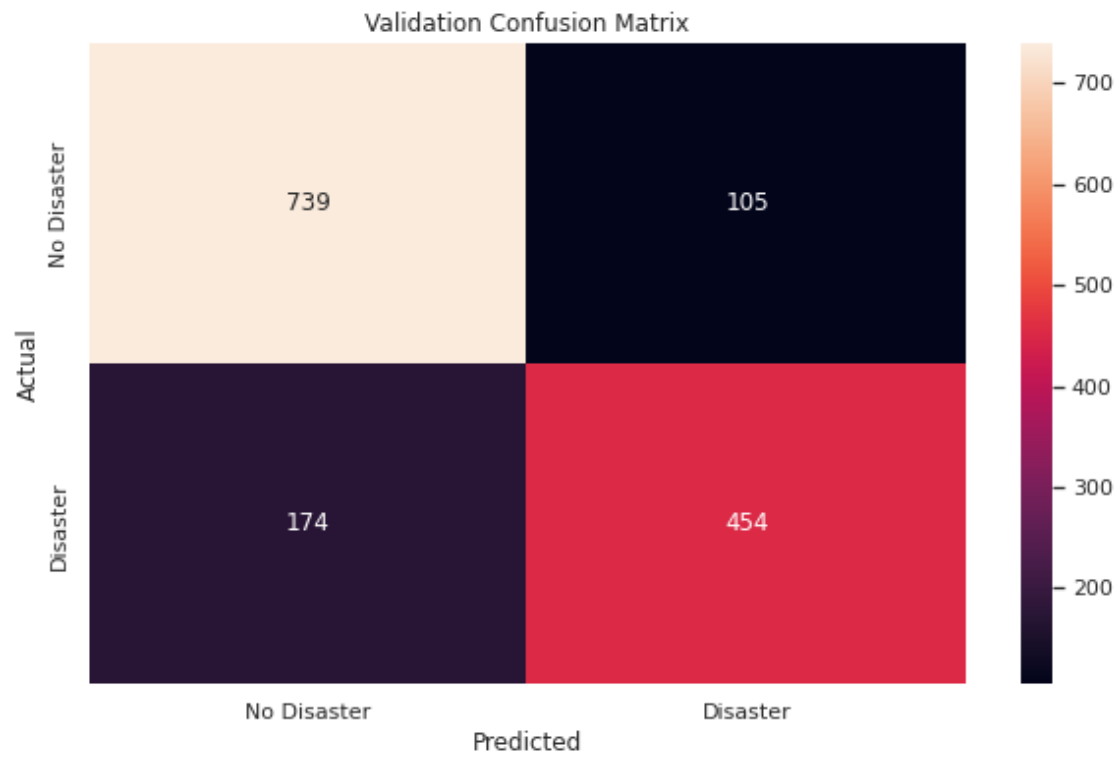
```
ax[1].plot(history.history['accuracy'])
ax[1].plot(history.history['val_accuracy'])
plot_history(history_bert)
```



```
[ ]: val_target = np.asarray([i[1] for i in list(val_ds.unbatch().
    ↪as_numpy_iterator())])
val_predict = bert_classifier.predict(val_ds)
fpr, tpr, _ = roc_curve(val_target, val_predict)
plt.plot(fpr, tpr, color='orange')
plt.plot([0,1], [0,1], linestyle='--')
plt.title('Validation ROC Curve')
plt.xlabel('False Positives (%)')
plt.ylabel('True Positives (%)')
plt.grid(True)
plt.show()
```



```
[ ]: THRESHOLD = 0.5 # Default value
# Get the true negative, false positive, false negative, and true positive values
tn, fp, fn, tp = confusion_matrix(val_target, val_predict > THRESHOLD).flatten()
# Construct the dataframe
cm = pd.DataFrame(
    [[tn, fp], [fn, tp]],
    index=['No Disaster', 'Disaster'],
    columns=['No Disaster', 'Disaster']
)
# Plot the matrix
sns.heatmap(cm, annot=True, fmt='g')
plt.title('Validation Confusion Matrix')
plt.ylabel('Actual')
plt.xlabel('Predicted')
plt.show()
```



[]: