aai-520-final-project-workflow

October 23, 2023

0.0.1 Install Python Libraries

```
[55]: !pip install kaggle
      !pip install tensorflow
      !pip install transformers
      !pip install evaluate
      !pip install scikit-learn
      !pip install datasets
      !pip install seaborn
      !pip install matplotlib
      !pip install typing
      !pip install pandas
```

```
Requirement already satisfied: kaggle in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (1.5.16)
Requirement already satisfied: six>=1.10 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (from kaggle) (1.16.0)
Requirement already satisfied: certifi in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (from kaggle) (2023.7.22)
Requirement already satisfied: python-dateutil in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (from kaggle) (2.8.2)
Requirement already satisfied: requests in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (from kaggle) (2.31.0)
Requirement already satisfied: tqdm in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (from kaggle) (4.65.0)
Requirement already satisfied: python-slugify in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (from kaggle) (8.0.1)
Requirement already satisfied: urllib3 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (from kaggle) (1.26.16)
Requirement already satisfied: bleach in
```

/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from kaggle) (6.1.0)
Requirement already satisfied: webencodings in /Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from bleach->kaggle) (0.5.1)
Requirement already satisfied: text-unidecode>=1.3 in /Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from python-slugify->kaggle) (1.3)
Requirement already satisfied: charset-normalizer<4,>=2 in /Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from requests->kaggle) (2.0.4)
Requirement already satisfied: idna<4,>=2.5 in /Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from requests->kaggle) (3.4)
Requirement already satisfied: tensorflow in /Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (2.12.0)
Requirement already satisfied: absl-py>=1.0.0 in /Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from tensorflow) (1.4.0)
Requirement already satisfied: astunparse>=1.6.0 in /Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from tensorflow) (1.6.3)
Requirement already satisfied: flatbuffers>=2.0 in /Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from tensorflow) (2.0)
Requirement already satisfied: gast<=0.4.0,>=0.2.1 in /Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from tensorflow) (0.4.0)
Requirement already satisfied: google-pasta>=0.1.1 in /Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from tensorflow) (0.2.0)
Requirement already satisfied: h5py>=2.9.0 in /Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from tensorflow) (3.7.0)
Requirement already satisfied: jax>=0.3.15 in /Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from tensorflow) (0.4.18)
Requirement already satisfied: libclang>=13.0.0 in /Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from tensorflow) (16.0.6)
Requirement already satisfied: numpy<1.24,>=1.22 in /Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from tensorflow) (1.23.5)
Requirement already satisfied: opt-einsum>=2.3.2 in /Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from tensorflow) (3.3.0)
Requirement already satisfied: packaging in

/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from tensorflow) (23.0)
Requirement already satisfied:
protobuf!=4.21.0,!=4.21.1,!=4.21.2,!=4.21.3,!=4.21.4,!=4.21.5,<5.0.0dev,>=3.20.3
in /Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from tensorflow) (3.20.3)
Requirement already satisfied: setuptools in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from tensorflow) (68.0.0)
Requirement already satisfied: six>=1.12.0 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from tensorflow) (1.16.0)
Requirement already satisfied: termcolor>=1.1.0 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from tensorflow) (2.1.0)
Requirement already satisfied: typing-extensions>=3.6.6 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from tensorflow) (4.7.1)
Requirement already satisfied: wrapt<1.15,>=1.11.0 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from tensorflow) (1.14.1)
Requirement already satisfied: grpcio<2.0,>=1.24.3 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from tensorflow) (1.48.2)
Requirement already satisfied: tensorboard<2.13,>=2.12 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from tensorflow) (2.12.1)
Requirement already satisfied: tensorflow-estimator<2.13,>=2.12.0 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from tensorflow) (2.12.0)
Requirement already satisfied: keras<2.13,>=2.12.0 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from tensorflow) (2.12.0)
Requirement already satisfied: wheel<1.0,>=0.23.0 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from astunparse>=1.6.0->tensorflow) (0.35.1)
Requirement already satisfied: ml-dtypes>=0.2.0 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from jax>=0.3.15->tensorflow) (0.3.1)
Requirement already satisfied: scipy>=1.7 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from jax>=0.3.15->tensorflow) (1.10.1)
Requirement already satisfied: google-auth<3,>=1.6.3 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from tensorboard<2.13,>=2.12->tensorflow) (2.6.0)
Requirement already satisfied: google-auth-oauthlib<1.1,>=0.5 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from tensorboard<2.13,>=2.12->tensorflow) (0.5.2)

Requirement already satisfied: markdown>=2.6.8 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (from tensorboard<2.13,>=2.12->tensorflow) (3.4.1)
Requirement already satisfied: requests<3,>=2.21.0 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (from tensorboard<2.13,>=2.12->tensorflow) (2.31.0)
Requirement already satisfied: tensorboard-data-server<0.8.0,>=0.7.0 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (from tensorboard<2.13,>=2.12->tensorflow) (0.7.0)
Requirement already satisfied: tensorboard-plugin-wit>=1.6.0 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (from tensorboard<2.13,>=2.12->tensorflow) (1.8.1)
Requirement already satisfied: werkzeug>=1.0.1 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (from tensorboard<2.13,>=2.12->tensorflow) (2.2.3)
Requirement already satisfied: cachetools<6.0,>=2.0.0 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (from google-auth<3,>=1.6.3->tensorboard<2.13,>=2.12->tensorflow)
(4.2.2)
Requirement already satisfied: pyasn1-modules>=0.2.1 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (from google-auth<3,>=1.6.3->tensorboard<2.13,>=2.12->tensorflow)
(0.2.8)
Requirement already satisfied: rsa<5,>=3.1.4 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (from google-auth<3,>=1.6.3->tensorboard<2.13,>=2.12->tensorflow)
(4.7.2)
Requirement already satisfied: requests-oauthlib>=0.7.0 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (from google-auth-
oauthlib<1.1,>=0.5->tensorboard<2.13,>=2.12->tensorflow) (1.3.0)
Requirement already satisfied: charset-normalizer<4,>=2 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (from requests<3,>=2.21.0->tensorboard<2.13,>=2.12->tensorflow) (2.0.4)
Requirement already satisfied: idna<4,>=2.5 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (from requests<3,>=2.21.0->tensorboard<2.13,>=2.12->tensorflow) (3.4)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (from requests<3,>=2.21.0->tensorboard<2.13,>=2.12->tensorflow)
(1.26.16)
Requirement already satisfied: certifi>=2017.4.17 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (from requests<3,>=2.21.0->tensorboard<2.13,>=2.12->tensorflow)
(2023.7.22)
Requirement already satisfied: MarkupSafe>=2.1.1 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (from werkzeug>=1.0.1->tensorboard<2.13,>=2.12->tensorflow) (2.1.1)

```
Requirement already satisfied: pyasn1<0.5.0,>=0.4.6 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (from pyasn1-modules>=0.2.1->google-
auth<3,>=1.6.3->tensorboard<2.13,>=2.12->tensorflow) (0.4.8)
Requirement already satisfied: oauthlib>=3.0.0 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (from requests-oauthlib>=0.7.0->google-auth-
oauthlib<1.1,>=0.5->tensorboard<2.13,>=2.12->tensorflow) (3.2.2)
Requirement already satisfied: transformers in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (4.32.1)
Requirement already satisfied: filelock in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (from transformers) (3.9.0)
Requirement already satisfied: huggingface-hub<1.0,>=0.15.1 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (from transformers) (0.15.1)
Requirement already satisfied: numpy>=1.17 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (from transformers) (1.23.5)
Requirement already satisfied: packaging>=20.0 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (from transformers) (23.0)
Requirement already satisfied: pyyaml>=5.1 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (from transformers) (6.0)
Requirement already satisfied: regex!=2019.12.17 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (from transformers) (2022.7.9)
Requirement already satisfied: requests in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (from transformers) (2.31.0)
Requirement already satisfied: tokenizers!=0.11.3,<0.14,>=0.11.1 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (from transformers) (0.13.2)
Requirement already satisfied: safetensors>=0.3.1 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (from transformers) (0.3.2)
Requirement already satisfied: tqdm>=4.27 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (from transformers) (4.65.0)
Requirement already satisfied: fsspec in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (from huggingface-hub<1.0,>=0.15.1->transformers) (2023.9.2)
Requirement already satisfied: typing-extensions>=3.7.4.3 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (from huggingface-hub<1.0,>=0.15.1->transformers) (4.7.1)
Requirement already satisfied: charset-normalizer<4,>=2 in
```

/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from requests->transformers) (2.0.4)
Requirement already satisfied: idna<4,>=2.5 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from requests->transformers) (3.4)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from requests->transformers) (1.26.16)
Requirement already satisfied: certifi>=2017.4.17 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from requests->transformers) (2023.7.22)
Requirement already satisfied: evaluate in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (0.4.0)
Requirement already satisfied: datasets>=2.0.0 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from evaluate) (2.12.0)
Requirement already satisfied: numpy>=1.17 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from evaluate) (1.23.5)
Requirement already satisfied: dill in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from evaluate) (0.3.6)
Requirement already satisfied: pandas in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from evaluate) (1.5.3)
Requirement already satisfied: requests>=2.19.0 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from evaluate) (2.31.0)
Requirement already satisfied: tqdm>=4.62.1 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from evaluate) (4.65.0)
Requirement already satisfied: xxhash in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from evaluate) (2.0.2)
Requirement already satisfied: multiprocess in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from evaluate) (0.70.14)
Requirement already satisfied: fsspec[http]>=2021.05.0 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from evaluate) (2023.9.2)
Requirement already satisfied: huggingface-hub>=0.7.0 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from evaluate) (0.15.1)
Requirement already satisfied: packaging in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from evaluate) (23.0)
Requirement already satisfied: responses<0.19 in

/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from evaluate) (0.13.3)
Requirement already satisfied: pyarrow>=8.0.0 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from datasets>=2.0.0->evaluate) (11.0.0)
Requirement already satisfied: aiohttp in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from datasets>=2.0.0->evaluate) (3.8.3)
Requirement already satisfied: pyyaml>=5.1 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from datasets>=2.0.0->evaluate) (6.0)
Requirement already satisfied: filelock in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from huggingface-hub>=0.7.0->evaluate) (3.9.0)
Requirement already satisfied: typing-extensions>=3.7.4.3 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from huggingface-hub>=0.7.0->evaluate) (4.7.1)
Requirement already satisfied: charset-normalizer<4,>=2 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from requests>=2.19.0->evaluate) (2.0.4)
Requirement already satisfied: idna<4,>=2.5 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from requests>=2.19.0->evaluate) (3.4)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from requests>=2.19.0->evaluate) (1.26.16)
Requirement already satisfied: certifi>=2017.4.17 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from requests>=2.19.0->evaluate) (2023.7.22)
Requirement already satisfied: six in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from responses<0.19->evaluate) (1.16.0)
Requirement already satisfied: python-dateutil>=2.8.1 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from pandas->evaluate) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from pandas->evaluate) (2022.7)
Requirement already satisfied: attrs>=17.3.0 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from aiohttp->datasets>=2.0.0->evaluate) (22.1.0)
Requirement already satisfied: multidict<7.0,>=4.5 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from aiohttp->datasets>=2.0.0->evaluate) (6.0.2)
Requirement already satisfied: async-timeout<5.0,>=4.0.0a3 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from aiohttp->datasets>=2.0.0->evaluate) (4.0.2)
Requirement already satisfied: yarl<2.0,>=1.0 in

/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from aiohttp->datasets>=2.0.0->evaluate) (1.8.1)
Requirement already satisfied: frozenlist>=1.1.1 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from aiohttp->datasets>=2.0.0->evaluate) (1.3.3)
Requirement already satisfied: aiosignal>=1.1.2 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from aiohttp->datasets>=2.0.0->evaluate) (1.2.0)
Requirement already satisfied: scikit-learn in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (1.2.2)
Requirement already satisfied: numpy>=1.17.3 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from scikit-learn) (1.23.5)
Requirement already satisfied: scipy>=1.3.2 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from scikit-learn) (1.10.1)
Requirement already satisfied: joblib>=1.1.1 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from scikit-learn) (1.2.0)
Requirement already satisfied: threadpoolctl>=2.0.0 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from scikit-learn) (2.2.0)
Requirement already satisfied: datasets in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (2.12.0)
Requirement already satisfied: numpy>=1.17 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from datasets) (1.23.5)
Requirement already satisfied: pyarrow>=8.0.0 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from datasets) (11.0.0)
Requirement already satisfied: dill<0.3.7,>=0.3.0 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from datasets) (0.3.6)
Requirement already satisfied: pandas in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from datasets) (1.5.3)
Requirement already satisfied: requests>=2.19.0 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from datasets) (2.31.0)
Requirement already satisfied: tqdm>=4.62.1 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from datasets) (4.65.0)
Requirement already satisfied: xxhash in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from datasets) (2.0.2)
Requirement already satisfied: multiprocess in

/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from datasets) (0.70.14)
Requirement already satisfied: fsspec[http]>=2021.11.1 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from datasets) (2023.9.2)
Requirement already satisfied: aiohttp in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from datasets) (3.8.3)
Requirement already satisfied: huggingface-hub<1.0.0,>=0.11.0 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from datasets) (0.15.1)
Requirement already satisfied: packaging in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from datasets) (23.0)
Requirement already satisfied: responses<0.19 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from datasets) (0.13.3)
Requirement already satisfied: pyyaml>=5.1 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from datasets) (6.0)
Requirement already satisfied: attrs>=17.3.0 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from aiohttp->datasets) (22.1.0)
Requirement already satisfied: charset-normalizer<3.0,>=2.0 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from aiohttp->datasets) (2.0.4)
Requirement already satisfied: multidict<7.0,>=4.5 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from aiohttp->datasets) (6.0.2)
Requirement already satisfied: async-timeout<5.0,>=4.0.0a3 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from aiohttp->datasets) (4.0.2)
Requirement already satisfied: yarl<2.0,>=1.0 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from aiohttp->datasets) (1.8.1)
Requirement already satisfied: frozenlist>=1.1.1 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from aiohttp->datasets) (1.3.3)
Requirement already satisfied: aiosignal>=1.1.2 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from aiohttp->datasets) (1.2.0)
Requirement already satisfied: filelock in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from huggingface-hub<1.0.0,>=0.11.0->datasets) (3.9.0)
Requirement already satisfied: typing-extensions>=3.7.4.3 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from huggingface-hub<1.0.0,>=0.11.0->datasets) (4.7.1)
Requirement already satisfied: idna<4,>=2.5 in

/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from requests>=2.19.0->datasets) (3.4)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from requests>=2.19.0->datasets) (1.26.16)
Requirement already satisfied: certifi>=2017.4.17 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from requests>=2.19.0->datasets) (2023.7.22)
Requirement already satisfied: six in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from responses<0.19->datasets) (1.16.0)
Requirement already satisfied: python-dateutil>=2.8.1 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from pandas->datasets) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from pandas->datasets) (2022.7)
Requirement already satisfied: seaborn in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (0.12.2)
Requirement already satisfied: numpy!=1.24.0,>=1.17 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from seaborn) (1.23.5)
Requirement already satisfied: pandas>=0.25 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from seaborn) (1.5.3)
Requirement already satisfied: matplotlib!=3.6.1,>=3.1 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from seaborn) (3.7.1)
Requirement already satisfied: contourpy>=1.0.1 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (1.0.5)
Requirement already satisfied: cycler>=0.10 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (4.25.0)
Requirement already satisfied: kiwisolver>=1.0.1 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (1.4.4)
Requirement already satisfied: packaging>=20.0 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (23.0)
Requirement already satisfied: pillow>=6.2.0 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (9.4.0)
Requirement already satisfied: pyparsing>=2.3.1 in

/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from pandas>=0.25->seaborn) (2022.7)
Requirement already satisfied: six>=1.5 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from python-dateutil>=2.7->matplotlib!=3.6.1,>=3.1->seaborn) (1.16.0)
Requirement already satisfied: matplotlib in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (3.7.1)
Requirement already satisfied: contourpy>=1.0.1 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from matplotlib) (1.0.5)
Requirement already satisfied: cycler>=0.10 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from matplotlib) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from matplotlib) (4.25.0)
Requirement already satisfied: kiwisolver>=1.0.1 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from matplotlib) (1.4.4)
Requirement already satisfied: numpy>=1.20 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from matplotlib) (1.23.5)
Requirement already satisfied: packaging>=20.0 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from matplotlib) (23.0)
Requirement already satisfied: pillow>=6.2.0 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from matplotlib) (9.4.0)
Requirement already satisfied: pyparsing>=2.3.1 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from matplotlib) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from matplotlib) (2.8.2)
Requirement already satisfied: six>=1.5 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (from python-dateutil>=2.7->matplotlib) (1.16.0)
Requirement already satisfied: typing in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-packages (3.7.4.3)
Requirement already satisfied: pandas in

```
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (1.5.3)
Requirement already satisfied: python-dateutil>=2.8.1 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (from pandas) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (from pandas) (2022.7)
Requirement already satisfied: numpy>=1.21.0 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (from pandas) (1.23.5)
Requirement already satisfied: six>=1.5 in
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages (from python-dateutil>=2.8.1->pandas) (1.16.0)
```

0.0.2 Apply Kaggle Configuration

0.0.3 Import Libraries

```python
[58]: import numpy as np # linear algebra
      import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
      import os
      import math


      import kaggle
      import tensorflow as tf
      from datasets import Dataset, concatenate_datasets
      from typing import Union, Optional

      from transformers import AutoTokenizer, AdamWeightDecay, BlenderbotTokenizer,␣
       ↪TFBlenderbotForConditionalGeneration,␣
       ↪DataCollatorForLanguageModeling,DataCollatorForSeq2Seq,DataCollatorWithPadding,␣
       ↪create_optimizer, TFAutoModelForCausalLM,TFAutoModelForSeq2SeqLM, pipeline
      # Used for model metric call when applying the metrics during training and␣
       ↪evaluation
      from transformers.keras_callbacks import KerasMetricCallback
      from tensorflow.keras.optimizers import Adam, Adagrad, SGD

      #from dataclasses import dataclass
      from transformers.tokenization_utils_base import PreTrainedTokenizerBase,␣
       ↪PaddingStrategy

      import evaluate

      # Import sklearn metrics
      from sklearn.metrics import accuracy_score, precision_score, recall_score,␣
       ↪f1_score, confusion_matrix
```

0.0.4 Import Dataset

- huggingface ubuntu dialog corpus
- https://huggingface.co/datasets/ubuntu_dialogs_corpus
- authenticate with Kaggle
- download the files from Kaggle
- add them to dataset module

```python
[59]: kaggle.api.authenticate()
```

Warning: Your Kaggle API key is readable by other users on this system! To fix
this, you can run 'chmod 600 /Users/hudsonshimanyula/.kaggle/kaggle.json'

```python
[61]: # Define the directory where the files are located
      directory = "./Ubuntu-dialogue-corpus"
```

```
[62]: # Define the file names and corresponding dataframe names
      files_dataframes = {
          "dialogueText_301.csv": "train_ubuntu_dialogue_df",
          "dialogueText_196.csv": "validation_ubuntu_dialogue_df",
          "dialogueText.csv": "test_ubuntu_dialogue_df"
      }
```

```
[63]: # Iterate through the files in the directory
      for filename in os.listdir(directory):
          # If the file is one of the files we're interested in
          if filename in files_dataframes:
              # Construct the full file path
              file_path = os.path.join(directory, filename)
              # Load the file into a dataframe and assign it to a variable with the↵
       ↪corresponding name
              globals()[files_dataframes[filename]] = pd.read_csv(file_path)
```

```
[64]: train_ubuntu_dialogue_df.head()
```

```
[64]:    folder dialogueID                    date       from        to  \
      0     301      1.tsv  2004-11-23T11:49:00.000Z  stuNNed       NaN
      1     301      1.tsv  2004-11-23T11:49:00.000Z  crimsun  stuNNed
      2     301      1.tsv  2004-11-23T11:49:00.000Z  stuNNed  crimsun
      3     301      1.tsv  2004-11-23T11:49:00.000Z  crimsun  stuNNed
      4     301      1.tsv  2004-11-23T11:50:00.000Z  stuNNed  crimsun

                                                    text
      0    any ideas why java plugin takes so long to load?
      1                                           java 1.4?
      2                                                 yes
      3                         java 1.5 loads _much_ faster
      4  noneus: how can i get 1.5 is there a .deb some…
```

```
[65]: validation_ubuntu_dialogue_df.head()
```

```
[65]:    folder dialogueID                    date       from        to  \
      0     301      1.tsv  2004-11-23T11:49:00.000Z  stuNNed       NaN
      1     301      1.tsv  2004-11-23T11:49:00.000Z  crimsun  stuNNed
      2     301      1.tsv  2004-11-23T11:49:00.000Z  stuNNed  crimsun
      3     301      1.tsv  2004-11-23T11:49:00.000Z  crimsun  stuNNed
      4     301      1.tsv  2004-11-23T11:50:00.000Z  stuNNed  crimsun

                                                    text
      0    any ideas why java plugin takes so long to load?
      1                                           java 1.4?
      2                                                 yes
      3                         java 1.5 loads _much_ faster
```

```
4  noneus: how can i get 1.5 is there a .deb some…
```

[66]:
```
# Review the downloaded dataset
test_ubuntu_dialogue_df.head()
```

[66]:
```
   folder  dialogueID                       date       from         to  \
0       3  126125.tsv  2008-04-23T14:55:00.000Z  bad_image        NaN
1       3  126125.tsv  2008-04-23T14:56:00.000Z  bad_image        NaN
2       3  126125.tsv  2008-04-23T14:57:00.000Z  lordleemo  bad_image
3       3   64545.tsv  2009-08-01T06:22:00.000Z   mechtech        NaN
4       3   64545.tsv  2009-08-01T06:22:00.000Z   mechtech        NaN


                                                 text
0  Hello folks, please help me a bit with the fol…
1  Did I choose a bad channel? I ask because you …
2  the second sentence is better english    and we…
3                                        Sock Puppe?t
4                                                WTF?
```

0.0.5 Exploratory Data Analysis

- get dataset info
- get dataset description

[67]:
```
train_ubuntu_dialogue_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16587830 entries, 0 to 16587829
Data columns (total 6 columns):
 #   Column      Dtype
---  ------      -----
 0   folder      int64
 1   dialogueID  object
 2   date        object
 3   from        object
 4   to          object
 5   text        object
dtypes: int64(1), object(5)
memory usage: 759.3+ MB
```

[68]:
```
validation_ubuntu_dialogue_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9212877 entries, 0 to 9212876
Data columns (total 6 columns):
 #   Column      Dtype
---  ------      -----
 0   folder      int64
```

```
 1    dialogueID   object
 2    date         object
 3    from         object
 4    to           object
 5    text         object
dtypes: int64(1), object(5)
memory usage: 421.7+ MB
```

[69]: `test_ubuntu_dialogue_df.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1038324 entries, 0 to 1038323
Data columns (total 6 columns):
 #    Column       Non-Null Count    Dtype
---   ------       --------------    -----
 0    folder       1038324 non-null  int64
 1    dialogueID   1038324 non-null  object
 2    date         1038324 non-null  object
 3    from         1038311 non-null  object
 4    to           566035 non-null   object
 5    text         1038237 non-null  object
dtypes: int64(1), object(5)
memory usage: 47.5+ MB
```

[70]: `train_ubuntu_dialogue_df['text'].describe()`

[70]:
```
count      16586581
unique     12520971
top             yes
freq          83685
Name: text, dtype: object
```

[71]: `validation_ubuntu_dialogue_df['text'].describe()`

[71]:
```
count      9212063
unique     7232452
top            yes
freq         47000
Name: text, dtype: object
```

[72]: `test_ubuntu_dialogue_df['text'].describe()`

[72]:
```
count      1038237
unique      863907
top         thanks
freq          6256
Name: text, dtype: object
```

0.0.6 Preprocess the datasets

- convert the dataframes to datasets
- initialize the tokenizer by selecting a pretrained model
- initialize the model by selecting a pretrained model
- initialize the datacollator
- apply tokenizer to datacollator
- tokenize the dataset with the tokenizer function
- build a tokenized dataset
- build function to split concatenated sequences into shorter chunks defined by block size and use that to build the label

```
[73]: train_ubuntu_dialogue_ds = Dataset.from_pandas(train_ubuntu_dialogue_df[0:10])
      validation_ubuntu_dialogue_ds = Dataset.
       ↪from_pandas(validation_ubuntu_dialogue_df[0:10])
      test_ubuntu_dialogue_ds = Dataset.from_pandas(test_ubuntu_dialogue_df[0:10])
```

```
[74]: train_ubuntu_dialogue_ds
```

```
[74]: Dataset({
          features: ['folder', 'dialogueID', 'date', 'from', 'to', 'text'],
          num_rows: 10
      })
```

```
[75]: tokenizer = BlenderbotTokenizer.from_pretrained('facebook/
       ↪blenderbot-400M-distill') #AutoTokenizer.from_pretrained('facebook/
       ↪blenderbot-400M-distill') #,truncation=True,padding="longest",␣
       ↪max_length=512, return_overflowing_tokens=True,return_offsets_mapping=True)␣
       ↪#,truncation=True
```

```
Downloading (…)olve/main/vocab.json: 100%|        | 127k/127k [00:00<00:00,
3.60MB/s]
Downloading (…)olve/main/merges.txt: 100%|        | 62.9k/62.9k [00:00<00:00,
75.5MB/s]
Downloading (…)okenizer_config.json: 100%|        | 1.15k/1.15k [00:00<00:00,
865kB/s]
Downloading (…)in/added_tokens.json: 100%|        | 16.0/16.0 [00:00<00:00,
57.1kB/s]
Downloading (…)cial_tokens_map.json: 100%|        | 772/772 [00:00<00:00,
5.11MB/s]
Downloading (…)lve/main/config.json: 100%|        | 1.57k/1.57k [00:00<00:00,
4.03MB/s]
```

```
[76]: tokenizer.pad_token = tokenizer.eos_token
```

```
[77]: # used for causal inference
      data_collator = DataCollatorForSeq2Seq(tokenizer=tokenizer, return_tensors='tf')
```

```
[78]: tokenizer
```

```
[78]: BlenderbotTokenizer(name_or_path='facebook/blenderbot-400M-distill',
      vocab_size=8008, model_max_length=128, is_fast=False, padding_side='right',
      truncation_side='right', special_tokens={'bos_token': AddedToken("<s>",
      rstrip=False, lstrip=False, single_word=False, normalized=True), 'eos_token':
      AddedToken("</s>", rstrip=False, lstrip=False, single_word=False,
      normalized=True), 'unk_token': AddedToken("<unk>", rstrip=False, lstrip=False,
      single_word=False, normalized=True), 'sep_token': AddedToken("</s>",
      rstrip=False, lstrip=False, single_word=False, normalized=True), 'pad_token':
      '</s>', 'cls_token': AddedToken("<s>", rstrip=False, lstrip=False,
      single_word=False, normalized=True), 'mask_token': AddedToken("<mask>",
      rstrip=False, lstrip=True, single_word=False, normalized=True)},
      clean_up_tokenization_spaces=True)
```

```python
[79]: def preprocess_function(examples):
          """
          This function preprocesses the input examples for the model training. It␣
      ↪tokenizes the input text, removes examples where input_ids is None, and␣
      ↪creates decoder_input_ids and labels.

          Args:
          examples (dict): A dictionary containing the input text.

          Returns:
          dict: A dictionary containing the tokenized input text, decoder_input_ids,␣
      ↪and labels.
          """
          # Tokenize the examples
          tokenized_examples = tokenizer([" ".join(x) for x in examples['text']],␣
      ↪truncation=True, padding="longest", max_length=512)

          # Remove examples where input_ids is None
          tokenized_examples = {k: v for k, v in tokenized_examples.items() if v is␣
      ↪not None}

          # Create decoder_input_ids and labels
          tokenized_examples["decoder_input_ids"] = [[tokenizer.bos_token_id] + ids +␣
      ↪[tokenizer.eos_token_id] for ids in tokenized_examples["input_ids"]]
          tokenized_examples["labels"] = [[tokenizer.bos_token_id] + ids + [tokenizer.
      ↪eos_token_id] for ids in tokenized_examples["input_ids"]]

          return tokenized_examples
```

```python
[80]: tokenized_train_ubuntu_dialogue_ds = train_ubuntu_dialogue_ds.map(
          preprocess_function,
          batched=True,
```

```
        num_proc=1,
        remove_columns=train_ubuntu_dialogue_ds.column_names,
    )
```

Map: 0%| | 0/10 [00:00<?, ? examples/s]

```
[81]: tokenized_validation_ubuntu_dialogue_ds = validation_ubuntu_dialogue_ds.map(
          preprocess_function,
          batched=True,
          num_proc=1,
          remove_columns=validation_ubuntu_dialogue_ds.column_names,
      )
```

```
[82]: tokenized_test_ubuntu_dialogue_ds = test_ubuntu_dialogue_ds.map(
          preprocess_function,
          batched=True,
          num_proc=1,
          remove_columns=test_ubuntu_dialogue_ds.column_names,
      )
```

Map: 0%| | 0/10 [00:00<?, ? examples/s]

```
[83]: tokenized_train_ubuntu_dialogue_ds
```

```
[83]: Dataset({
          features: ['input_ids', 'attention_mask', 'decoder_input_ids', 'labels'],
          num_rows: 10
      })
```

0.0.7 Model Build

- build the hyperparameters
- build and compute metrics for model evaluation
- build the pretrained model
- build the tensorflow prepared dataset (includes tensorflow tensor conversion, batch setting for model runs, shuffling the data, adding the data collator for collation)
- build metric callbacks
- apply them to the model
- compile the model

```
[85]: model = TFBlenderbotForConditionalGeneration.from_pretrained('facebook/
      ↪blenderbot-400M-distill')
```

Downloading tf_model.h5: 100%| | 1.46G/1.46G [00:28<00:00, 51.8MB/s]
All model checkpoint layers were used when initializing
TFBlenderbotForConditionalGeneration.

19

[86]:
```python
# Preparing TensorFlow dataset for training:
# The `prepare_tf_dataset` method is called on the model object to convert the
 ↪tokenized training dataset into a TensorFlow dataset.
# The dataset is shuffled to ensure randomness, which is beneficial for
 ↪training the model.
# Batch size is set to 1, meaning that the dataset will be divided into
 ↪batches, each containing 1 example.
# The `collate_fn` function is passed to handle the batching and padding of
 ↪sequences.
tf_train_set = model.prepare_tf_dataset(
    tokenized_train_ubuntu_dialogue_ds,
    shuffle=True,
    batch_size=1,
    collate_fn=data_collator,
)


# Preparing TensorFlow dataset for validation:
# Similar to the training set preparation, but the dataset is not shuffled in
 ↪this case.
# This dataset will be used for validating the model's performance during
 ↪training.
tf_validation_set = model.prepare_tf_dataset(
    tokenized_validation_ubuntu_dialogue_ds,
    shuffle=False,
    batch_size=1,
    collate_fn=data_collator,
)


# Preparing TensorFlow dataset for testing:
# Similar to the validation set preparation, this dataset will be used for
 ↪evaluating the model's performance after training.
tf_test_set = model.prepare_tf_dataset(
    tokenized_test_ubuntu_dialogue_ds,
    shuffle=False,
    batch_size=1,
    collate_fn=data_collator,
)
```

```
[89]:  # Define a list of optimizers
       optimizers = ['adam','sgd', 'adagrad']
```

```
[90]:  # Parameters for capturing the best optimizer used in the model
       best_eval_loss = float('inf')
       best_model = None
       best_optimizer = None
```

0.0.8 Model Compile, Train and Evaluate

- Iterate over the optimizers to determine the best one to be used
- compile the model
- train the model on the train and validation datasets
- evaluate the model based on loss leader of optimizer

```
[93]:  # Iterate over the optimizers
       for optimizer in optimizers:
           if optimizer == 'adam':
               opt = tf.keras.optimizers.Adam(learning_rate=2e-5, weight_decay=0.01) #␣
       ↪decay = weight_decay
           elif optimizer == 'adafactor':
               opt = tf.keras.optimizers.Adafactor(learning_rate=2e-5, weight_decay=0.
       ↪01)
           elif optimizer == 'sgd':
               opt = tf.keras.optimizers.SGD(learning_rate=2e-5, weight_decay=0.01)
           elif optimizer == 'adagrad':
               opt = tf.keras.optimizers.Adagrad(learning_rate=2e-5, weight_decay=0.01)

           model.compile(optimizer=opt) # can also add metrics here

           # Train the model
           model.fit(x=tf_train_set, validation_data=tf_validation_set, epochs=3) #␣
       ↪callbacks=callbacks

           # Evaluate the model
           eval_loss = model.evaluate(tf_validation_set)  # callbacks=callbacks

           # If this model has a lower evaluation loss than the current best, update␣
       ↪the best model and optimizer
           if eval_loss < best_eval_loss:
               best_eval_loss = eval_loss
               best_model = model
               best_optimizer = optimizer
```

WARNING:absl:At this time, the v2.11+ optimizer `tf.keras.optimizers.Adam` runs
slowly on M1/M2 Macs, please use the legacy Keras optimizer instead, located at
`tf.keras.optimizers.legacy.Adam`.

```
WARNING:absl:There is a known slowdown when using v2.11+ Keras optimizers on
M1/M2 Macs. Falling back to the legacy Keras optimizer, i.e.,
`tf.keras.optimizers.legacy.Adam`.

Epoch 1/3

2023-10-15 23:05:50.762160: W
tensorflow/tsl/platform/profile_utils/cpu_utils.cc:128] Failed to get CPU
frequency: 0 Hz

10/10 [==============================] - 25s 1s/step - loss: 1.1745 - val_loss:
0.4351
Epoch 2/3
10/10 [==============================] - 9s 966ms/step - loss: 0.3026 -
val_loss: 0.1261
Epoch 3/3
10/10 [==============================] - 9s 976ms/step - loss: 0.0799 -
val_loss: 0.0415
10/10 [==============================] - 3s 294ms/step - loss: 0.0415

WARNING:absl:At this time, the v2.11+ optimizer `tf.keras.optimizers.SGD` runs
slowly on M1/M2 Macs, please use the legacy Keras optimizer instead, located at
`tf.keras.optimizers.legacy.SGD`.
WARNING:absl:There is a known slowdown when using v2.11+ Keras optimizers on
M1/M2 Macs. Falling back to the legacy Keras optimizer, i.e.,
`tf.keras.optimizers.legacy.SGD`.

Epoch 1/3
10/10 [==============================] - 22s 2s/step - loss: 0.0454 - val_loss:
0.0413
Epoch 2/3
10/10 [==============================] - 10s 1000ms/step - loss: 0.0515 -
val_loss: 0.0412
Epoch 3/3
10/10 [==============================] - 10s 1s/step - loss: 0.0465 - val_loss:
0.0411
10/10 [==============================] - 3s 276ms/step - loss: 0.0411

WARNING:absl:At this time, the v2.11+ optimizer `tf.keras.optimizers.Adagrad`
runs slowly on M1/M2 Macs, please use the legacy Keras optimizer instead,
located at `tf.keras.optimizers.legacy.Adagrad`.
WARNING:absl:There is a known slowdown when using v2.11+ Keras optimizers on
M1/M2 Macs. Falling back to the legacy Keras optimizer, i.e.,
`tf.keras.optimizers.legacy.Adagrad`.

Epoch 1/3
10/10 [==============================] - 23s 1s/step - loss: 0.0462 - val_loss:
0.0407
Epoch 2/3
10/10 [==============================] - 10s 1s/step - loss: 0.0453 - val_loss:
0.0403
```

```
Epoch 3/3
10/10 [==============================] - 9s 957ms/step - loss: 0.0429 -
val_loss: 0.0399
10/10 [==============================] - 2s 239ms/step - loss: 0.0399
```

[94]:
```python
print('Best Model to use based on the optimizer: ', best_model)
print('Best Optimizer to use based loss leader comparison: ', best_optimizer)
```

```
Best Model to use based on the optimizer:  <transformers.models.blenderbot.model
ing_tf_blenderbot.TFBlenderbotForConditionalGeneration object at 0x5990eb0d0>
Best Optimizer to use based loss leader comparison:  adagrad
```

0.0.9 Model Evaluate

- evaluate the model's performance using the model evaluate and model predict

[95]:
```python
eval_results = best_model.evaluate(tf_validation_set)
print(f'Perplexity: {math.exp(eval_results):.2f}')
```

```
10/10 [==============================] - 2s 233ms/step - loss: 0.0399
Perplexity: 1.04
```

[134]:
```python
acc = evaluate.load('accuracy')
f1_score = evaluate.load('f1')
precis = evaluate.load('precision')
recalll = evaluate.load('recall')
```

[97]:
```python
# Round the predictions to turn them into "0" or "1" labels
test_preds = model.predict(tf_validation_set)
```

```
10/10 [==============================] - 6s 232ms/step
```

[116]:
```python
predicted_ids = tf.argmax(test_preds.logits, axis=-1)
predicted_ids_tensor = tf.constant(predicted_ids, dtype=tf.int64)
references = tf.constant(tokenized_validation_ubuntu_dialogue_ds['labels'],
  ↪dtype=tf.int64)

predicted_ids_numpy = predicted_ids_tensor.numpy().flatten()
references_numpy = references.numpy().flatten()
```

[113]:
```python
accuracy_score = acc.compute(predictions=predicted_ids_numpy,
  ↪references=references_numpy)

print('Accuracy Score: ', accuracy_score)
```

```
Accuracy Score:  {'accuracy': 0.9907692307692307}
```

[ ]:

```python
[135]:  # Calculating precision score:
        # The `compute` method is called on the precision object (`precis`) to
        ↪calculate the precision score.
        # Predictions and references are converted to numpy arrays and flattened before
        ↪being passed to the `compute` method.
        # The 'weighted' average is used to calculate a single precision score that
        ↪takes class imbalance into account.
        precision_score = precis.compute(predictions=predicted_ids_tensor.numpy().
        ↪flatten(), references=references.numpy().flatten(), average='weighted')


        # Calculating recall score:
        # Similarly, the `compute` method is called on the recall object (`recalll`) to
        ↪calculate the recall score.
        # The 'weighted' average is used to calculate a single recall score that takes
        ↪class imbalance into account.
        recall_score = recalll.compute(predictions=predicted_ids_tensor.numpy().
        ↪flatten(), references=references.numpy().flatten(), average='weighted')


        # Calculating F1 score:
        # The `compute` method is called on the F1 score object (`f1_score`) to
        ↪calculate the F1 score.
        # The 'weighted' average is used to calculate a single F1 score that takes
        ↪class imbalance into account.
        f1_score = f1_score.compute(predictions=predicted_ids_tensor.numpy().flatten(),
        ↪references=references.numpy().flatten(), average='weighted')


        # Printing the calculated scores:
        print('Precision Score: ', precision_score)
        print('Recall Score: ', recall_score)
        print('F1 Score: ', f1_score)
```

```
precision Score:  {'precision': 0.9865765004226542}
recall Score:  {'recall': 0.9907692307692307}
F1 Score:  {'f1': 0.9878009561060409}

/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages/sklearn/metrics/_classification.py:1344: UndefinedMetricWarning:
Precision is ill-defined and being set to 0.0 in labels with no predicted
samples. Use `zero_division` parameter to control this behavior.
  _warn_prf(average, modifier, msg_start, len(result))
/Users/hudsonshimanyula/anaconda3/envs/AI_MASTERS_ENV/lib/python3.11/site-
packages/sklearn/metrics/_classification.py:1344: UndefinedMetricWarning: Recall
is ill-defined and being set to 0.0 in labels with no true samples. Use
`zero_division` parameter to control this behavior.
  _warn_prf(average, modifier, msg_start, len(result))
```

0.0.10 Model Save

```
[136]: # Define the directory where the files are located
       save_model_directory = "./bestModel"
```

```
[137]: ### Save the weights of the model
       best_model.save_pretrained(save_model_directory)
```

0.0.11 Perform Inference

- recall the saved model
- check the model's architecture
- test the model to confirm it can perform causal inference calculations properly

```
[5]: from transformers import BlenderbotTokenizer,␣
      ↪TFBlenderbotForConditionalGeneration
     import tensorflow as tf

     # Load the model and tokenizer from the directory where you saved your model
     save_model_directory = "./bestModel"
     tokenizer = BlenderbotTokenizer.from_pretrained('facebook/
      ↪blenderbot-400M-distill')
     model= TFBlenderbotForConditionalGeneration.
      ↪from_pretrained(save_model_directory)

     def get_response(user_input):
         # Tokenize the user input
         inputs = tokenizer(user_input, return_tensors='tf').input_ids
         # Generate a response
         outputs = model.generate(input_ids=inputs, max_new_tokens=100,␣
      ↪do_sample=True, top_k=30, top_p=0.85)
         # Decode the response
         response = tokenizer.batch_decode(outputs, skip_special_tokens=True)[0]
         return response


     def print_welcome_message():
         print("=======================================")
         print("    Welcome to the Chatbot Interface!    ")
         print(" Type 'quit', 'exit', or 'bye' to exit ")
         print("=======================================")

     print_welcome_message()

     # Interactive loop
     while True:
         user_input = input("You: ")
```

```python
    if user_input.lower() in ['quit', 'exit', 'bye']:
        print("Bot: Bye! Have a great day.")
        break
    print(f"You: {user_input}")
    response = get_response(user_input)
    print(f"Bot: {response}")
```

All model checkpoint layers were used when initializing
TFBlenderbotForConditionalGeneration.

All the layers of TFBlenderbotForConditionalGeneration were initialized from the
model checkpoint at ./bestModel.
If your task is similar to the task the model of the checkpoint was trained on,
you can already use TFBlenderbotForConditionalGeneration for predictions without
further training.

```
========================================
   Welcome to the Chatbot Interface!
 Type 'quit', 'exit', or 'bye' to exit
========================================
You: Hello
Bot:  Hi, how are you doing today? I'm just relaxing before work, how about you?
You: where do you work
Bot:  I work at a grocery store. It's not the best but it pays the bills.
You: Well thats good. Straight to the point, I need help with my linux system
Bot:  I know how you feel.  I need to get a new one as well.
You:
Bot:  I'm not sure what you mean by that, but I hope you have a great day!
You: Can you help me troubleshoot an issue
Bot:  Sure, what is the issue? I can help you if you want to talk about it.
Bot: Bye! Have a great day.
```