```
!pip install transformers
```

```
Requirement already satisfied: transformers in /usr/local/lib/python3.6/dist-package
Requirement already satisfied: protobuf in /usr/local/lib/python3.6/dist-packages (f
Requirement already satisfied: packaging in /usr/local/lib/python3.6/dist-packages (
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.6/dist-pa
Requirement already satisfied: sentencepiece!=0.1.92 in /usr/local/lib/python3.6/dis
Requirement already satisfied: requests in /usr/local/lib/python3.6/dist-packages (f
Requirement already satisfied: dataclasses; python_version < "3.7" in /usr/local/lib
Requirement already satisfied: tqdm>=4.27 in /usr/local/lib/python3.6/dist-packages
Requirement already satisfied: sacremoses in /usr/local/lib/python3.6/dist-packages
Requirement already satisfied: filelock in /usr/local/lib/python3.6/dist-packages (f
Requirement already satisfied: numpy in /usr/local/lib/python3.6/dist-packages (from
Requirement already satisfied: tokenizers==0.9.2 in /usr/local/lib/python3.6/dist-pa
Requirement already satisfied: setuptools in /usr/local/lib/python3.6/dist-packages
Requirement already satisfied: six>=1.9 in /usr/local/lib/python3.6/dist-packages (f
Requirement already satisfied: pyparsing>=2.0.2 in /usr/local/lib/python3.6/dist-pac
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.6/dist-p
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.6/dist-package
Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in /usr/local
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.6/dist-pa
Requirement already satisfied: joblib in /usr/local/lib/python3.6/dist-packages (fro
Requirement already satisfied: click in /usr/local/lib/python3.6/dist-packages (from
```

```
import transformers
transformers.logging.set_verbosity_error()
```

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

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.
```

```
! rm -r *.py
! cp -r '/content/drive/My Drive/tweet-sentiment-extraction/py_files/.' '/content'
```

```
import pandas as pd
test_df = pd.read_csv('/content/drive/My Drive/tweet-sentiment-extraction/test.csv')
test_df.shape
```

```
(3534, 3)
```

```
train_df = pd.read_csv('/content/drive/My Drive/tweet-sentiment-extraction/train.csv')
print(train_df.shape)
```

```
(27481, 4)
```

```
from preprocess_data import preprocess_data
from model import create_model
from predict_output import predict
from get_metric import get_metric
from tokenizer import get_tokenizer
```

Function1

- Take raw data as input and return Predictions for that point

```
def get_predictions(input):
  tokenizer = get_tokenizer('/content/drive/My Drive/tweet-sentiment-extraction/roberta_to
  input_ids,attention_mask,input = preprocess_data(input,128,tokenizer)
  model = create_model(128,0.1,'/content/drive/My Drive/tweet-sentiment-extraction/mymodel
  input_data = (input_ids,attention_mask)
  pred_text= predict(model,input_data,tokenizer,input)
  return pred_text
```

```
ip = test_df.sample(10)
op = get_predictions(ip)
from prettytable import PrettyTable
myTable = PrettyTable(["Sentiment", "Input text", "Output text"])
for i in range(len(op)):
  myTable.add_row([ip['sentiment'].values[i],ip['text'].values[i],op[i]])

print(myTable)
```

```
10it [00:00, 269.72it/s]Loading Pretrained Tokenizer for TfRoberta model
************************************************
Preprocessing input data
Getting input_ids and attention_mask for the input
Shape of input id and attention mask: (10, 128) (10, 128)

************************************************
Loaded Pretrained TFRobertaForQuestionAnswering model
************************************************
Loaded trained model
************************************************
+-----------+---------------------------------------------------------------------------
| Sentiment |                                                                Input text
+-----------+---------------------------------------------------------------------------
|  negative |               i`m going to kill myself t_t . i wasted hundreds of down
|  positive |                   it was an awesome talk  find it very true that i am
|  positive |                                      can`t wait to hear the evp! and c
|  positive |                                                 happy mother`s day ever
|  positive |               is wanting someone to spend her summer evening with whil
|  negative |                                             omfg. one of the worst day
|  positive |                                              staying home because i`m
|  neutral  |                                        who to say hi to and who to buy a
|  negative |                                             i`m so very tired...and have
|  neutral  |   most name brands have dairy even if it`s called `semi-sweet` or `dar
+-----------+---------------------------------------------------------------------------
```

Function 2

- Take raw data as input and return performance metric

```
def get_performance_metric(input):
  #Use existing get_predictions function to get predicted_text
  output = get_predictions(input)
  score = get_metric(input,output)
  return score


ip = train_df.sample(10)
score = get_performance_metric(ip)
print('Mean Jaccard score for given data:',score)
```

```
Loading Pretrained Tokenizer for TfRoberta model
**************************************************
10it [00:00, 1008.32it/s]Preprocessing input data
Getting input_ids and attention_mask for the input
Shape of input id and attention mask: (10, 128) (10, 128)

**************************************************
Loaded Pretrained TFRobertaForQuestionAnswering model
**************************************************
Loaded trained model
**************************************************
Successfully ran!!!!!
**************************************************
Mean Jaccard score for given data: 0.7936898395721925
```