+ Code   + Text   RAM/Disk

```python
[3]  import tensorflow as tf
     import tensorflow_hub as hub
     import tensorflow_text as text
```

```python
!pip install tensorflow_text
```

```
Collecting tensorflow_text
  Downloading tensorflow_text-2.14.0-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (6.5 MB)
     6.5/6.5 MB 38.6 MB/s eta 0:00:00
Requirement already satisfied: tensorflow-hub>=0.13.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow_text) (0.15.0)
Requirement already satisfied: tensorflow<2.15,>=2.14.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow_text) (2.14.0)
Requirement already satisfied: absl-py>=1.0.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow<2.15,>=2.14.0->tensorflow_text) (1.4.0)
Requirement already satisfied: astunparse>=1.6.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow<2.15,>=2.14.0->tensorflow_text) (1.6.3)
Requirement already satisfied: flatbuffers>=23.5.26 in /usr/local/lib/python3.10/dist-packages (from tensorflow<2.15,>=2.14.0->tensorflow_text) (23.5.26)
Requirement already satisfied: gast>=0.5.0,!=0.5.1,!=0.5.2,>=0.2.1 in /usr/local/lib/python3.10/dist-packages (from tensorflow<2.15,>=2.14.0->tensorflow_text) (0.5.4)
Requirement already satisfied: google-pasta>=0.1.1 in /usr/local/lib/python3.10/dist-packages (from tensorflow<2.15,>=2.14.0->tensorflow_text) (0.2.0)
Requirement already satisfied: h5py>=2.9.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow<2.15,>=2.14.0->tensorflow_text) (3.9.0)
Requirement already satisfied: libclang>=13.0.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow<2.15,>=2.14.0->tensorflow_text) (16.0.6)
Requirement already satisfied: ml-dtypes==0.2.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow<2.15,>=2.14.0->tensorflow_text) (0.2.0)
Requirement already satisfied: numpy>=1.23.5 in /usr/local/lib/python3.10/dist-packages (from tensorflow<2.15,>=2.14.0->tensorflow_text) (1.23.5)
Requirement already satisfied: opt-einsum>=2.3.2 in /usr/local/lib/python3.10/dist-packages (from tensorflow<2.15,>=2.14.0->tensorflow_text) (3.3.0)
Requirement already satisfied: packaging in /usr/local/lib/python3.10/dist-packages (from tensorflow<2.15,>=2.14.0->tensorflow_text) (23.2)
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 /usr/local/lib/python3.10/dist-packages (from tensorflow<2.15,>=2.14.0->tensorflow_text) (3.20.3)
Requirement already satisfied: setuptools in /usr/local/lib/python3.10/dist-packages (from tensorflow<2.15,>=2.14.0->tensorflow_text) (67.7.2)
Requirement already satisfied: six>=1.12.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow<2.15,>=2.14.0->tensorflow_text) (1.16.0)
Requirement already satisfied: termcolor>=1.1.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow<2.15,>=2.14.0->tensorflow_text) (2.3.0)
Requirement already satisfied: typing-extensions>=3.6.6 in /usr/local/lib/python3.10/dist-packages (from tensorflow<2.15,>=2.14.0->tensorflow_text) (4.5.0)
Requirement already satisfied: wrapt<1.15,>=1.11.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow<2.15,>=2.14.0->tensorflow_text) (1.14.1)
Requirement already satisfied: tensorflow-io-gcs-filesystem>=0.23.1 in /usr/local/lib/python3.10/dist-packages (from tensorflow<2.15,>=2.14.0->tensorflow_text) (0.34.0)
Requirement already satisfied: grpcio<2.0,>=1.24.3 in /usr/local/lib/python3.10/dist-packages (from tensorflow<2.15,>=2.14.0->tensorflow_text) (1.59.2)
Requirement already satisfied: tensorboard<2.15,>=2.14 in /usr/local/lib/python3.10/dist-packages (from tensorflow<2.15,>=2.14.0->tensorflow_text) (2.14.1)
Requirement already satisfied: tensorflow-estimator<2.15,>=2.14.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow<2.15,>=2.14.0->tensorflow_text) (2.14.0)
Requirement already satisfied: keras<2.15,>=2.14.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow<2.15,>=2.14.0->tensorflow_text) (2.14.0)
Requirement already satisfied: wheel<1.0,>=0.23.0 in /usr/local/lib/python3.10/dist-packages (from astunparse>=1.6.0->tensorflow<2.15,>=2.14.0->tensorflow_text) (0.41.3)
Requirement already satisfied: google-auth<3,>=1.6.3 in /usr/local/lib/python3.10/dist-packages (from tensorboard<2.15,>=2.14->tensorflow<2.15,>=2.14.0->tensorflow_text) (2.17.3)
Requirement already satisfied: google-auth-oauthlib<1.1,>=0.5 in /usr/local/lib/python3.10/dist-packages (from tensorboard<2.15,>=2.14->tensorflow<2.15,>=2.14.0->tensorflow_text) (1.0.0)
Requirement already satisfied: markdown>=2.6.8 in /usr/local/lib/python3.10/dist-packages (from tensorboard<2.15,>=2.14->tensorflow<2.15,>=2.14.0->tensorflow_text) (3.5.1)
Requirement already satisfied: requests<3,>=2.21.0 in /usr/local/lib/python3.10/dist-packages (from tensorboard<2.15,>=2.14->tensorflow<2.15,>=2.14.0->tensorflow_text) (2.31.0)
Requirement already satisfied: tensorboard-data-server<0.8.0,>=0.7.0 in /usr/local/lib/python3.10/dist-packages (from tensorboard<2.15,>=2.14->tensorflow<2.15,>=2.14.0->tensorflow_text) (0.7.2)
Requirement already satisfied: werkzeug>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from tensorboard<2.15,>=2.14->tensorflow<2.15,>=2.14.0->tensorflow_text) (3.0.1)
Requirement already satisfied: cachetools<6.0,>=2.0.0 in /usr/local/lib/python3.10/dist-packages (from google-auth<3,>=1.6.3->tensorboard<2.15,>=2.14->tensorflow<2.15,>=2.14.0->tensorflow_text) (5.3.2)
Requirement already satisfied: pyasn1-modules>=0.2.1 in /usr/local/lib/python3.10/dist-packages (from google-auth<3,>=1.6.3->tensorboard<2.15,>=2.14->tensorflow<2.15,>=2.14.0->tensorflow_text) (0.3.0)
Requirement already satisfied: rsa<5,>=3.1.4 in /usr/local/lib/python3.10/dist-packages (from google-auth<3,>=1.6.3->tensorboard<2.15,>=2.14->tensorflow<2.15,>=2.14.0->tensorflow_text) (4.9)
Requirement already satisfied: requests-oauthlib>=0.7.0 in /usr/local/lib/python3.10/dist-packages (from google-auth-oauthlib<1.1,>=0.5->tensorboard<2.15,>=2.14->tensorflow<2.15,>=2.14.0->tensorflow_text) (1.3.1)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=2.21.0->tensorboard<2.15,>=2.14->tensorflow<2.15,>=2.14.0->tensorflow_text) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=2.21.0->tensorboard<2.15,>=2.14->tensorflow<2.15,>=2.14.0->tensorflow_text) (3.4)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=2.21.0->tensorboard<2.15,>=2.14->tensorflow<2.15,>=2.14.0->tensorflow_text) (2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=2.21.0->tensorboard<2.15,>=2.14->tensorflow<2.15,>=2.14.0->tensorflow_text) (2023.7.22)
Requirement already satisfied: MarkupSafe>=2.1.1 in /usr/local/lib/python3.10/dist-packages (from werkzeug>=1.0.1->tensorboard<2.15,>=2.14->tensorflow<2.15,>=2.14.0->tensorflow_text) (2.1.3)
Requirement already satisfied: pyasn1<0.6.0,>=0.4.6 in /usr/local/lib/python3.10/dist-packages (from pyasn1-modules>=0.2.1->google-auth<3,>=1.6.3->tensorboard<2.15,>=2.14->tensorflow<2.15,>=2.14.0->tensorflow_text) (0.5.0)
Requirement already satisfied: oauthlib>=3.0.0 in /usr/local/lib/python3.10/dist-packages (from requests-oauthlib>=0.7.0->google-auth-oauthlib<1.1,>=0.5->tensorboard<2.15,>=2.14->tensorflow<2.15,>=2.14.0->tensorflow_text) (3.2.2)
Installing collected packages: tensorflow_text
Successfully installed tensorflow_text-2.14.0
```

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

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

+ Code   + Text

```python
[5]  import pandas as pd

     df = pd.read_csv("/content/drive/My Drive/Colab Notebooks/Dataset/spam.csv")
     df.head(5)
```

	Category	Message
0	ham	Go until jurong point, crazy.. Available only ...
1	ham	Ok lar... Joking wif u oni...
2	spam	Free entry in 2 a wkly comp to win FA Cup fina...
3	ham	U dun say so early hor... U c already then say...
4	ham	Nah I don't think he goes to usf, he lives aro...

```python
[6]  df['Category'].value_counts()
```

```
ham     4825
spam     747
Name: Category, dtype: int64
```

```python
[7]  747/4825
```

```
0.15481865284974095
```

15% spam emails, 85% ham emails: This indicates class imbalance

```python
[8]  df_spam = df[df['Category']=='spam']
     df_spam.shape
```

```
(747, 2)
```

```python
[9]  df_ham = df[df['Category']=='ham']
     df_ham.shape
```

```
(4825, 2)
```

```python
[10]  df_ham_downsampled = df_ham.sample(df_spam.shape[0])
      df_ham_downsampled.shape
```

```
(747, 2)
```

```python
[11]  df_balanced = pd.concat([df_ham_downsampled, df_spam])
      df_balanced.shape
```

```
(1494, 2)
```

```python
[12]  df_balanced['Category'].value_counts()
```

```
ham     747
spam    747
Name: Category, dtype: int64
```

```python
[13]  df_balanced['spam']=df_balanced['Category'].apply(lambda x: 1 if x=='spam' else 0)
      df_balanced.sample(5)
```

	Category	Message	spam
5141	spam	FREE for 1st week! No1 Nokia tone 4 ur mobile ...	1
42	spam	07732584351 - Rodger Burns - MSG = We tried to...	1
3524	ham	Try neva mate!!	0
1910	ham	Then u going ikea str aft dat?	0
3106	ham	Or remind me in a few hrs.	0

Split it into training and test data set

```python
[14]  from sklearn.model_selection import train_test_split

      X_train, X_test, y_train, y_test = train_test_split(df_balanced['Message'],df_balanced['spam'], stratify=df_balanced['spam'])
```

Now lets import BERT model and get embeding vectors for few sample statements

```python
[15]  bert_preprocess = hub.KerasLayer("https://tfhub.dev/tensorflow/bert_en_uncased_preprocess/3")
      bert_encoder = hub.KerasLayer("https://tfhub.dev/tensorflow/bert_en_uncased_L-12_H-768_A-12/4")
```

Build Model

There are two types of models you can build in tensorflow.

(1) Sequential (2) Functional

So far we have built sequential model. But below we will build functional model. More information on these two is here: https://becominghuman.ai/sequential-vs-functional-model-in-keras-20684f766057

```python
[16]  # Bert layers
      text_input = tf.keras.layers.Input(shape=(), dtype=tf.string, name='text')
      preprocessed_text = bert_preprocess(text_input)
      outputs = bert_encoder(preprocessed_text)

      # Neural network layers
      l = tf.keras.layers.Dropout(0.1, name="dropout")(outputs['pooled_output'])
      l = tf.keras.layers.Dense(1, activation='sigmoid', name="output")(l)

      # Use inputs and outputs to construct a final model
      model = tf.keras.Model(inputs=[text_input], outputs = [l])
```

```python
[17]  model.summary()
```

```
Model: "model"
_____
Layer (type)                 Output Shape              Param #    Connected to
=========================================================================================
text (InputLayer)            [(None,)]                 0          []

keras_layer (KerasLayer)     {'input_mask': (None, 128) 0         ['text[0][0]']
                             , 'input_type_ids': (None,
                             128),
                             'input_word_ids': (None,
                             128)}

keras_layer_1 (KerasLayer)   {'encoder_outputs': [(None 1094822   ['keras_layer[0][0]',
                             , 128, 768),            41           'keras_layer[0][1]',
                             (None, 128, 768),                    'keras_layer[0][2]']
                             (None, 128, 768),
                             (None, 128, 768),
                             (None, 128, 768),
                             (None, 128, 768),
                             (None, 128, 768),
                             (None, 128, 768),
                             (None, 128, 768),
                             (None, 128, 768),
                             (None, 128, 768)],
                             'default': (None, 768),
                             'sequence_output': (None,
                             128, 768),
                             'pooled_output': (None, 7
                             68)}

dropout (Dropout)            (None, 768)               0          ['keras_layer_1[0][13]']

output (Dense)               (None, 1)                 769        ['dropout[0][0]']

=========================================================================================
Total params: 109483010 (417.64 MB)
Trainable params: 769 (3.00 KB)
Non-trainable params: 109482241 (417.64 MB)
```

```python
[19]  METRICS = [
              tf.keras.metrics.BinaryAccuracy(name='accuracy'),
              tf.keras.metrics.Precision(name='precision'),
              tf.keras.metrics.Recall(name='recall')
      ]

      model.compile(optimizer='adam',
                    loss='binary_crossentropy',
                    metrics=METRICS)
```

```python
model.fit(X_train, y_train, epochs=10)
```

```
Epoch 1/10
24/35 [===================>..........] - ETA: 3:06 - loss: 0.6717 - accuracy: 0.5768 - precision: 0.5867 - recall: 0.5273
```

```python
[ ]  import numpy as np

     y_predicted = np.where(y_predicted > 0.5, 1, 0)
     y_predicted
```

```python
sample_dataset = [
    'You can win alot of money, register in the link below',
    'You have an iphone 10, spin the image below to claim your prize and it willl be delivered in your door step',
    'You have an offer, the company will give you 50% off in every item purchased.',
    'Hey Bravin, dont be late for the meeting tomorrow, it will start at exactly 10:30am,
    "See you monday, we have alot to talk about the future of this company ."
]
model.predict(sample_dataset)
```