

TF pipeline 2

May 11, 2023

```
[1]: import tensorflow as tf

[2]: tf.__version__

[2]: '2.11.0'

[51]: reviews = tf.data.Dataset.list_files(r'C:\Users\Rakesh\Downloads\reviews\*', shuffle=False)

[52]: type(reviews)

[52]: tensorflow.python.data.ops.from_tensor_slices_op.TensorSliceDataset

[53]: for file in reviews:
        print(file.numpy())

b'C:\\Users\\Rakesh\\Downloads\\reviews\\negative\\neg_1.txt'
b'C:\\Users\\Rakesh\\Downloads\\reviews\\negative\\neg_2.txt'
b'C:\\Users\\Rakesh\\Downloads\\reviews\\negative\\neg_3.txt'
b'C:\\Users\\Rakesh\\Downloads\\reviews\\positive\\pos_1.txt'
b'C:\\Users\\Rakesh\\Downloads\\reviews\\positive\\pos_2.txt'
b'C:\\Users\\Rakesh\\Downloads\\reviews\\positive\\pos_3.txt'

[64]: import os
def extract_review_and_label(file_path):
    return tf.io.read_file(file_path), tf.strings.split(file_path, os.path.
    ↪sep)[-2]

    # do not try single for ex
```

1 Extract review text from these files. Extract label from folder name

```
[65]: reviews_1 = reviews.map(extract_review_and_label)
for review, label in reviews_1:
    print("Review: ", review.numpy()[:50])
    print("Label: ", label.numpy())
```

```

Review: b"Basically there's a family where a little boy (Jak"
Label: b'negative'
Review: b'This show was an amazing, fresh & innovative idea '
Label: b'negative'
Review: b''
Label: b'negative'
Review: b'One of the other reviewers has mentioned that afte'
Label: b'positive'
Review: b'A wonderful little production. <br /><br />The fil'
Label: b'positive'
Review: b''
Label: b'positive'

```

2 Filter blank reviews

```

[66]: reviews_2 = reviews_1.filter( lambda reviews ,label : reviews != '' )
      for review, label in reviews_2.as_numpy_iterator():
          print("Review: ",review[:50])
          print("Label: ",label)

```

```

WARNING:tensorflow:From C:\Users\Rakesh\anaconda3\lib\site-
packages\tensorflow\python\autograph\pyct\static_analysis\liveness.py:83:
Analyzer.lamba_check (from
tensorflow.python.autograph.pyct.static_analysis.liveness) is deprecated and
will be removed after 2023-09-23.
Instructions for updating:
Lambda fuctions will be no more assumed to be used in the statement where they
are used, or at least in the same block.
https://github.com/tensorflow/tensorflow/issues/56089
Review: b"Basically there's a family where a little boy (Jak"
Label: b'negative'
Review: b'This show was an amazing, fresh & innovative idea '
Label: b'negative'
Review: b'One of the other reviewers has mentioned that afte'
Label: b'positive'
Review: b'A wonderful little production. <br /><br />The fil'
Label: b'positive'

```

3 Perform map, filter and shuffle all in single line of code

```

[69]: one_line = reviews.map(extract_review_and_label).filter( lambda reviews ,label :
      ↪ reviews != '' ).shuffle(3)
      for review, label in reviews_2.as_numpy_iterator():
          print("Review: ",review[:50])
          print("Label: ",label)

```

```

Review: b"Basically there's a family where a little boy (Jak"

```

Label: b'negative'
Review: b'This show was an amazing, fresh & innovative idea '
Label: b'negative'
Review: b'One of the other reviewers has mentioned that afte'
Label: b'positive'
Review: b'A wonderful little production.

The fil'
Label: b'positive'

[]: