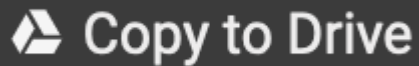


## ▼ Instructions

Please make a copy of this file before continuing. You can click on the "Copy to Drive" button above. It looks like this:



## ▼ 1 Setup

We first need to install `textprocessing` and `pysentimiento`. Loading `pysentimiento` will take a while so please be patient :)

```
!pip install git+https://github.com/geil002/textprocessing.git --no-cache-dir -c
```

```
    Preparing metadata (setup.py) ... done
    Building wheel for tp (setup.py) ... done
```

If installing `textprocessing` directly does not work, we will need to upgrade `setuptools`, `pip` and `'wheels'` first. Remove the `#` before running the code below, only if you encounter errors when installing `textprocessing`.

```
#!pip install --upgrade setuptools
#!pip install --upgrade pip
```

```
pip install -q pysentimiento
```

```

_____ 251.2/251.2 kB 12.2 MB/s eta
_____ 519.6/519.6 kB 42.6 MB/s eta
_____ 175.4/175.4 kB 16.1 MB/s eta
Preparing metadata (setup.py) ... done
_____ 670.2/670.2 MB 2.0 MB/s eta (
_____ 7.7/7.7 MB 39.2 MB/s eta 0:00
_____ 115.3/115.3 kB 10.3 MB/s eta
_____ 194.1/194.1 kB 20.1 MB/s eta
_____ 134.8/134.8 kB 9.9 MB/s eta (
_____ 302.0/302.0 kB 29.9 MB/s eta
_____ 23.7/23.7 MB 61.8 MB/s eta 0
_____ 823.6/823.6 kB 43.2 MB/s eta
_____ 14.1/14.1 MB 65.2 MB/s eta 0
_____ 731.7/731.7 MB 2.3 MB/s eta (
_____ 410.6/410.6 MB 2.6 MB/s eta (
_____ 121.6/121.6 MB 8.1 MB/s eta (
_____ 56.5/56.5 MB 12.7 MB/s eta 0
_____ 124.2/124.2 MB 8.1 MB/s eta (
_____ 196.0/196.0 MB 6.3 MB/s eta (
_____ 209.8/209.8 MB 5.6 MB/s eta (
```

```

===== 99.1/99.1 kB 10.4 MB/s eta 0
===== 89.2/89.2 MB 9.7 MB/s eta 0:0
===== 20.2/20.2 MB 58.7 MB/s eta 0
===== 3.8/3.8 MB 91.8 MB/s eta 0:0
===== 1.3/1.3 MB 22.3 MB/s eta 0:0
===== 295.0/295.0 kB 32.8 MB/s eta

```

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

```

ERROR: pip's dependency resolver does not currently take into account all
fastai 2.7.12 requires torch<2.1,>=1.7, but you have torch 2.1.0 which is
torchaudio 2.0.2+cud118 requires torch==2.0.1, but you have torch 2.1.0 whi
torchdata 0.6.1 requires torch==2.0.1, but you have torch 2.1.0 which is in
torchtext 0.15.2 requires torch==2.0.1, but you have torch 2.1.0 which is
torchvision 0.15.2+cud118 requires torch==2.0.1, but you have torch 2.1.0 w

```

## 2 Loading data

To see if this work, we can fetch the sample data seen in the tutorial.

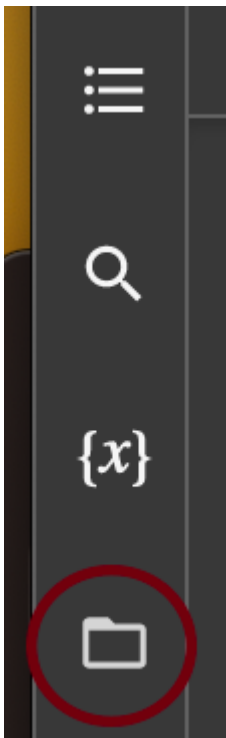
```

!wget -q https://github.com/geil002/textprocessing/raw/main/data.zip
!unzip -q data.zip

```

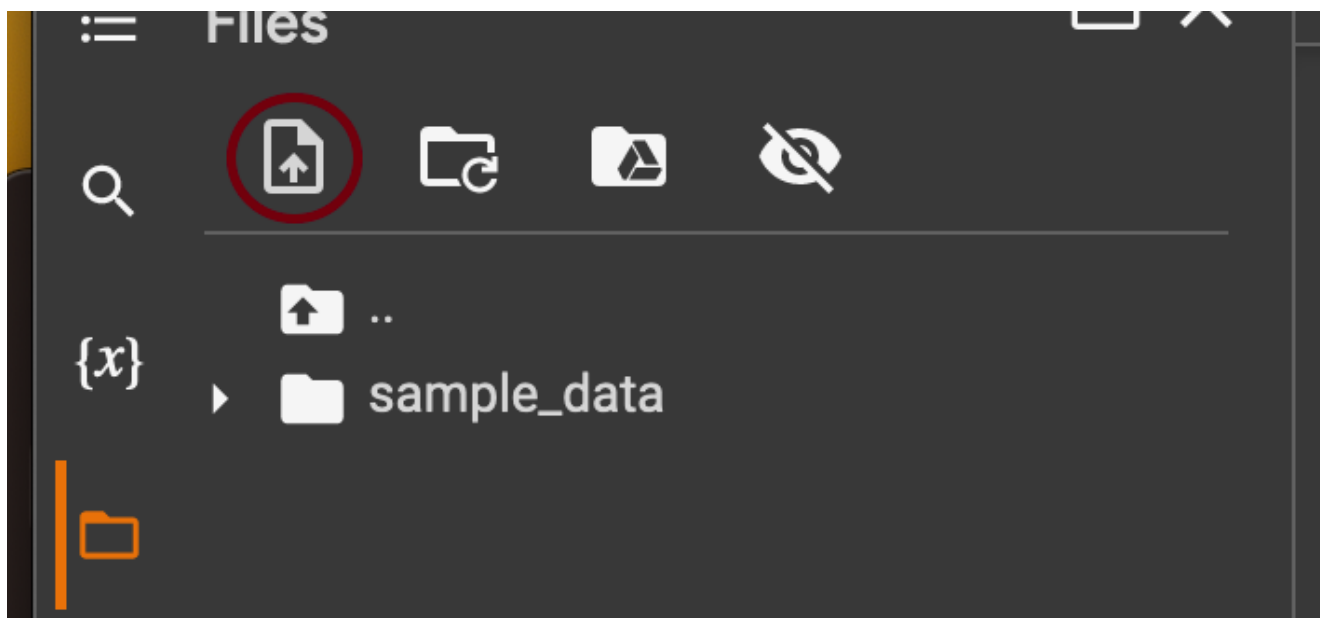
Now let's work with *our own* data. As stated in worksheet 3.3, each member of your group must label the reviews of our fictional cafe (coffee.xlsx) as 'pos', 'neg' or 'neu'. To upload this data to colab you must go to the the Files navigator in Colab, add a new folder and upload the files made by your groupmates.

1) Click on the Files button on the left of the notebook.

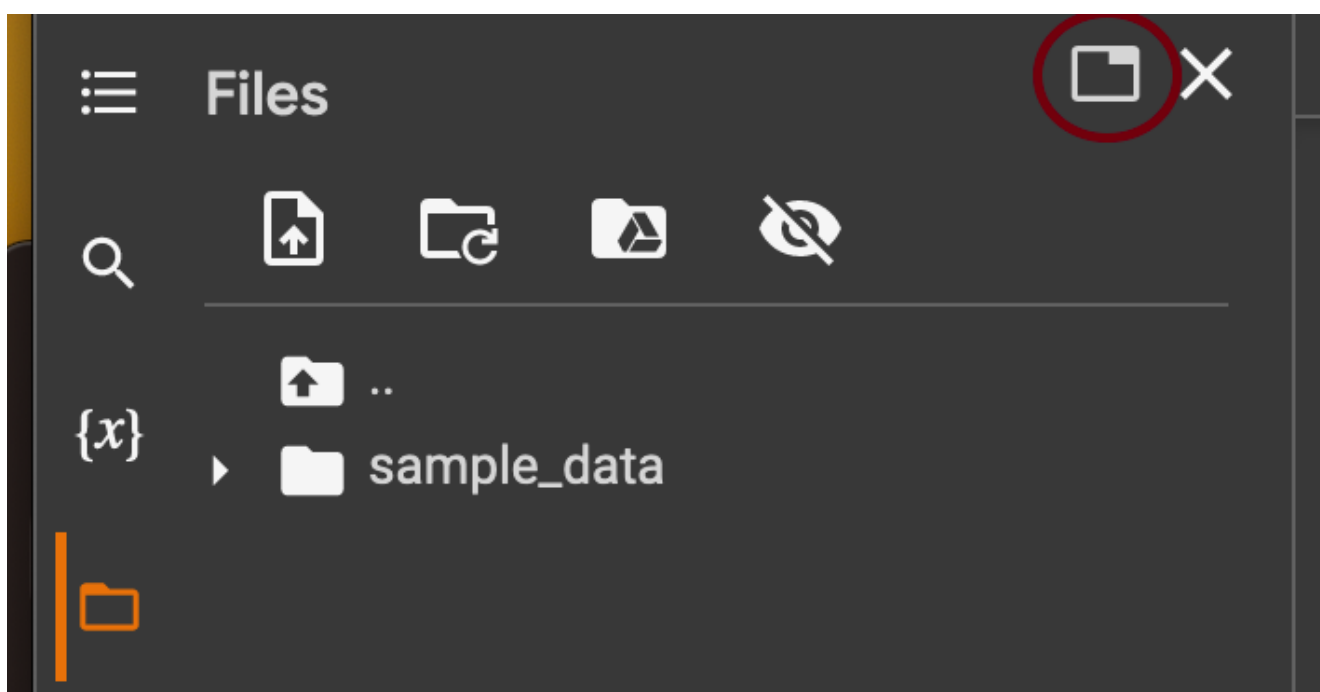


2) Click upload and select a file.





3) If you want to organize your uploads into subfolders, you can click on the "add a new folder" button. If this doesn't work, click with the right button of your mouse and select "new folder".



### 3 Loading modules

```
import textprocessing as tp
```

```
Downloading (...)lve/main
```

```
949/949 [00:00<00:00,
```

```
/config.json: 100%
```

```
32.1kB/s]
```

```
Downloading
```

```
540M/540M [00:03<00:00,
```

```
pytorch_model.bin: 100%
```

```
171MB/s]
```

```
Downloading 338/338 [00:00<00:00,
(...)okenizer_config.json: 100% 8.00kB/s]
Downloading (...)solve/main 843k/843k [00:00<00:00,
/vocab.txt: 100% 12.6MB/s]
```

## 4 Analysis

This is the same as what you saw in the videos for week 6.

```
tp.compare_raters("travel_blogs")
```

```
Processing travel_blogs/blogs_labeler1.xlsx...
Processing travel_blogs/blogs_labeler2.xlsx...
Processing travel_blogs/blogs_labeler3.xlsx...
```

```
Agreement file saved at travel_blogs_output/agreement.xlsx
Fliess Kappa is 0.6592389670263585
```

If your excel files don't have an index column, you can pass in the parameter `includes_index = False`

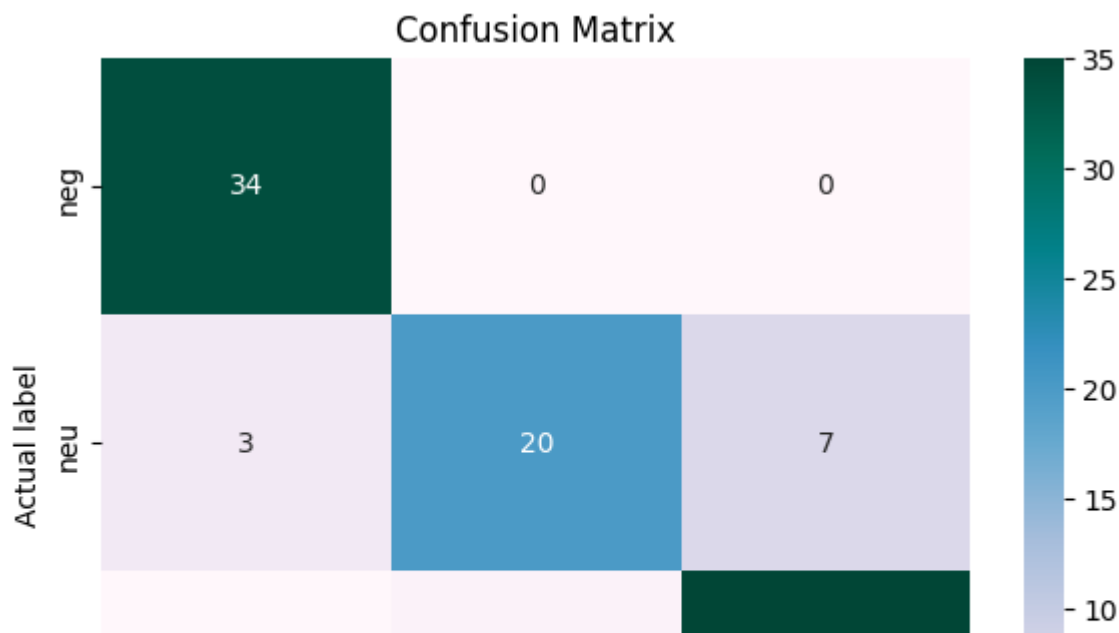
```
#tp.compare_raters("no_index",includes_index=False)
```

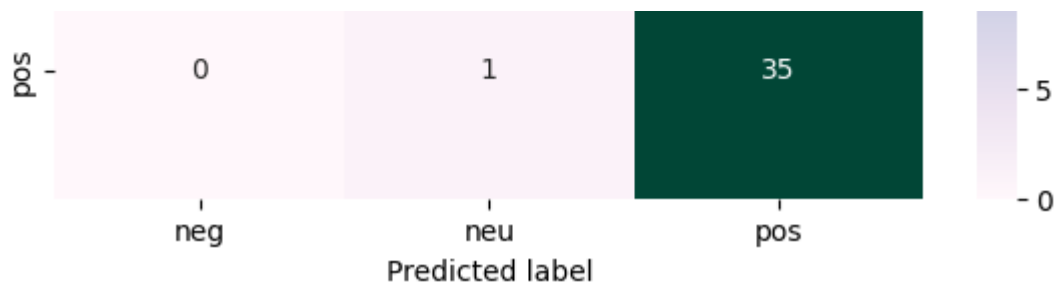
```
ERROR: There must be at least three files in the directory.
```

```
sa = tp.sentiment_analysis("travel_blogs_output/agreement.xlsx")
```

```
The accuracy score is 89.0
```

```
sa.confusion_matrix()
```





```
sa.examine_labels(actual="neu",predicted="neg")
```

3

|    | sentence                                  | blogs_labeler1 | blogs_labeler2 | blogs_labeler3 | ground truth | prec |
|----|---|----------------|----------------|----------------|--------------|------|
| 15 | A week in the countryside was uneventful. | neu            | neu            | neu            | neu          |      |
| 65 | My train journey to work was as usual,    | neu            | neu            | neu            | neu          |      |