

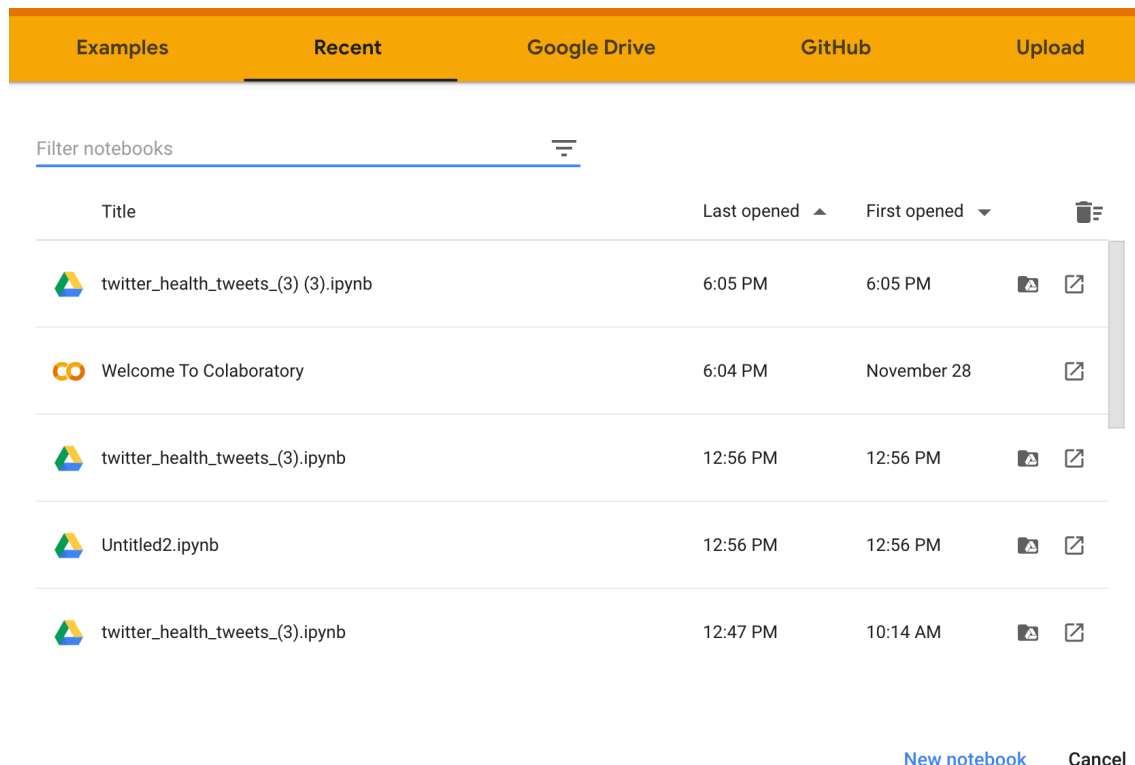
readME

Github Link: <https://github.com/NaveenAre04/Twitter-Health-Surveillance>

This contains three files and readMe file.

- 1) Source code file: twitter_health_tweets.ipynb
- 2) Test csv file: Corona_NLP_test.csv
- 3) Train csv file: Corona_NLP_train.csv

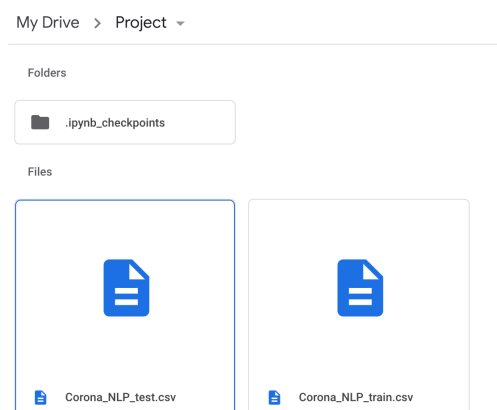
- 1) Open google colab in the google chrome.
- 2) Click on the “upload” button on displayed window.



- 3) Upload the source code file which we provide.

Code file name: twitter_health_tweets.ipynb

- 4) Create a folder named “Project” in your goole drive and please upload test and train datasets which we will provide you.



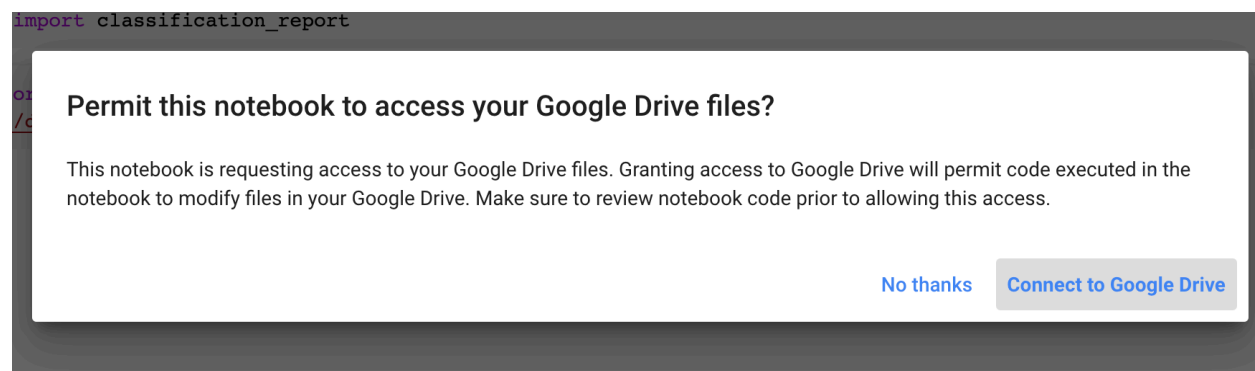
Test data: Corona_NLP_test.csv
Train data: Corona_NLP_train.csv

5) Comeback to google colab and execute first kernel which contains all the import packages of python and machine learning.

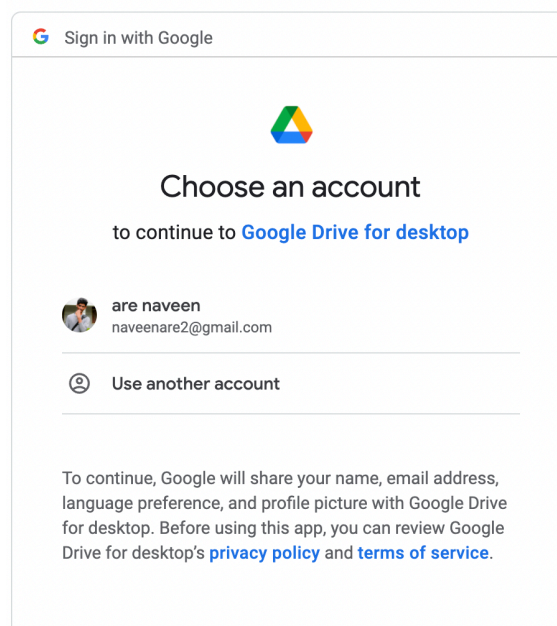
6) Next execute the kernel which is after to the packages. This is for mounting the drive and execute it.

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

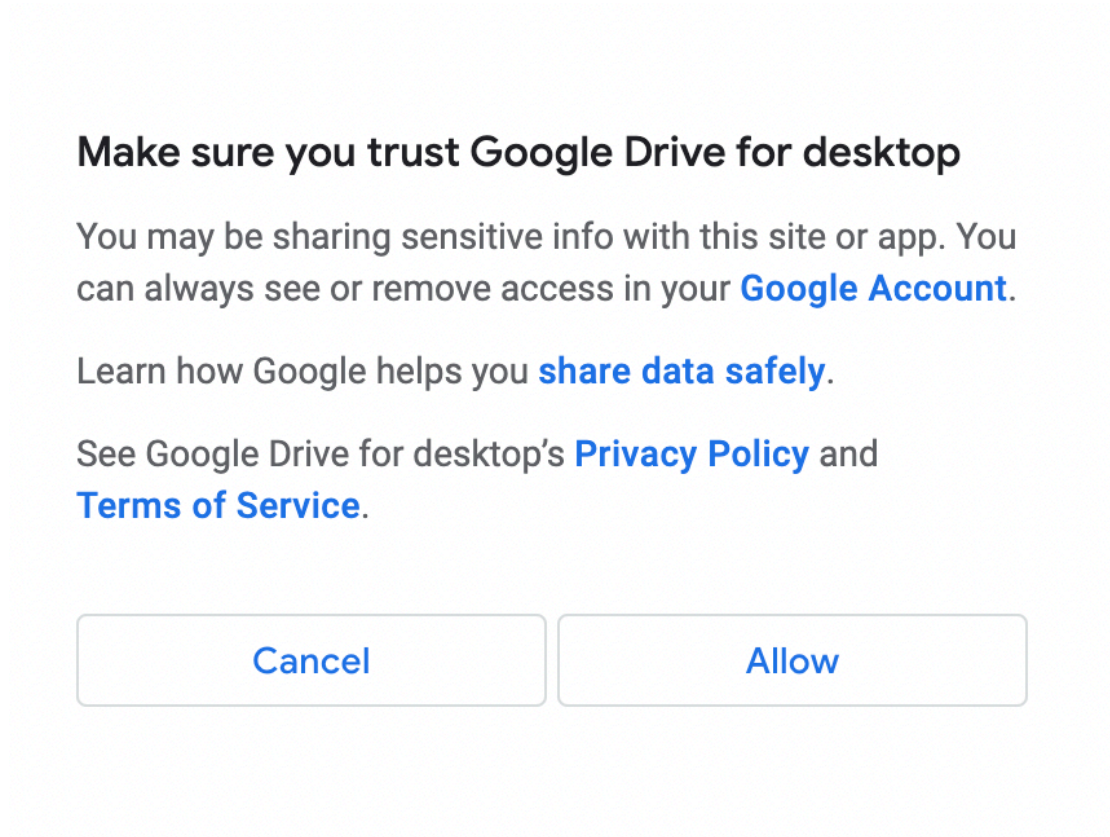
7) It will prompt for the permission to connect the google drive like below and click on connect to google drive.



8) Choose the google account in which you stored train and test data.csv files. The account must be linked to the drive in which we created the folder and placed our files.

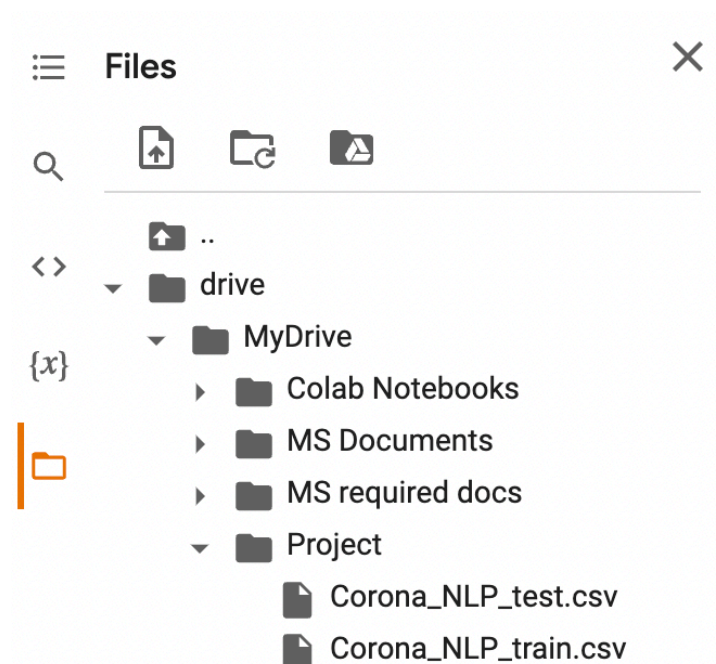


9) click on the allow button which will gives permission to access the files in google drive.



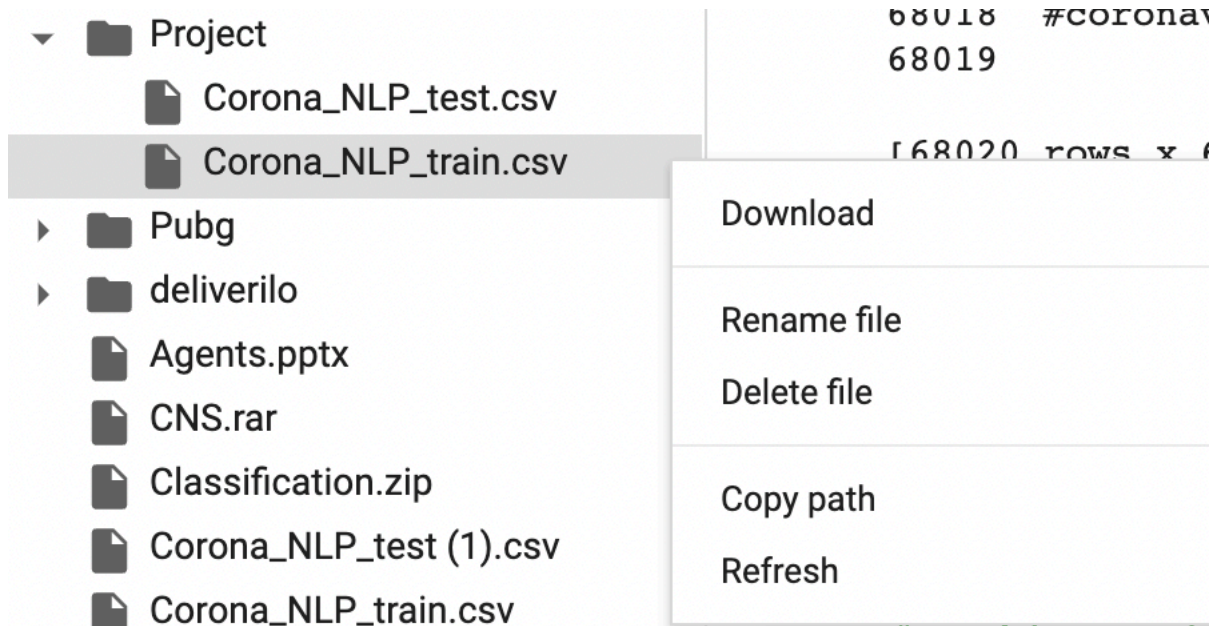
10)

Then in google colab where we opened our source code, under files section it will show our data sets files in the "Project" folder and this will be under MyDrive parent folder.



Note: Please execute remaining kernels in order.

11) Please copy the path of the files by clicking on “copy path” option whenever it requires



Note : Only these two kernels need paths, so please paste path exactly to avoid manipulations.

12) Below two kernels needs the path of those files. Please copy the appropriate path of those files and paste it in `pd.read_csv()`.

Kernel 1- Here For spark load, copy the `Corona_NLP_train.csv` path and paste it in `read_csv` function of `df1` variable. Below is the snapshot of `df1` variable.

```
df1 = spark.read.load("/content/drive/MyDrive/Project/Corona_NLP_train.csv", format="csv", inferSchema=True, header=True)
```

Kernel 2- Below are the snapshots for train and test path kernels.

For train and test please copy the appropriate file paths and paste it in the respective functions.

For train, copy path of `Corona_NLP_train.csv` file and paste it in `pd.read_csv()` function of train variable

For test, copy path of `Corona_NLP_test.csv` file and paste it in `pd.read_csv()` function of test variable.

```
train = pd.read_csv('/content/drive/MyDrive/Project/Corona_NLP_test.csv', encoding='latin1')
test = pd.read_csv('/content/drive/MyDrive/Project/Corona_NLP_test.csv', encoding='latin1')

# Combine train and test set
df2 = train.append(test, ignore_index=True)
df = df.append(test, ignore_index=True)
```

13) Please execute all the kernels in order, so that results will be displayed properly.