

How to solve the NLP challenge

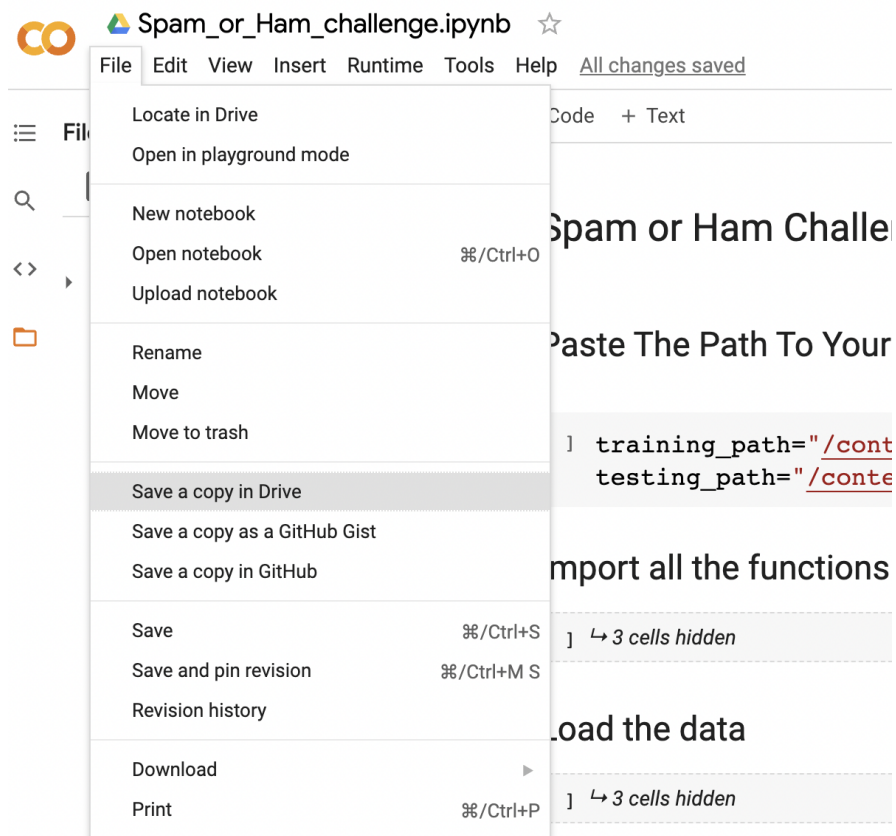
1- Download the datasets ([link](#) to datasets)

2- Open training dataset in excel and label texts as 'ham' or 'spam'

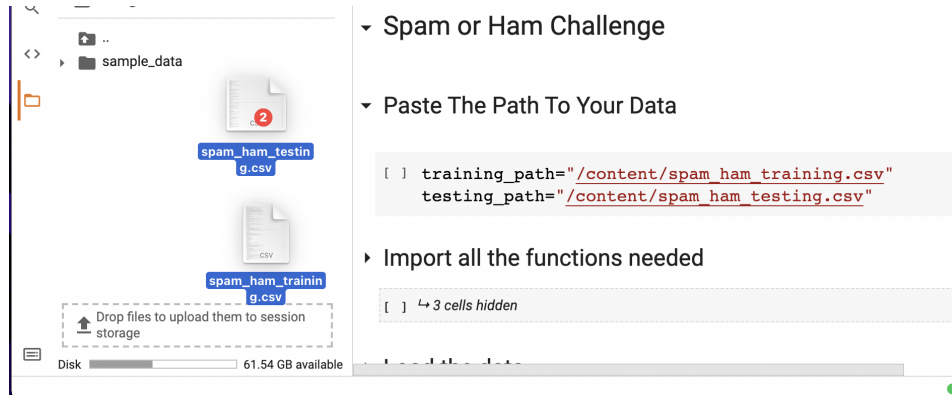
94	ham	Smile in Pleasure Smile in Pain Smile when trouble pours like Rain Smile when sun1 Hurts U Smile becoz SOMEONE :
95	spam	Please call our customer service representative on 0800 169 6031 between 10am-9pm as you have WON a guaranteee
96	ham	Havent planning to buy later. I check already lido only got 530 show in e afternoon. U finish work already?
97	spam	Your free ringtone is waiting to be collected. Simply text the password \MIX\" to 85069 to verify. C PO Box 5249
98	ham	Watching telugu movie..wat abt u?
99	ham	i see. When we finish we have loads of loans to pay
100	ham	Hi. Wk been ok - on hols now! Yes on for a bit of a run. Forgot that i have hairdressers appointment at four so need to
101		Please don't text me anymore. I have nothing else to say.
102		Okay name ur price as long as its legal! Wen can I pick them up? Y u ave x ams xx
103		I'm still looking for a car to buy. And have not gone 4the driving test yet.
104		As per your request 'Melle Melle (Oru Minnaminunginte Nurungu Vettam)' has been set as your callertune for all Calle
105		wow. You're right! I didn't mean to do that. I guess once i gave up on boston men and changed my search location to r
106		Umma my life and vava umma love you lot dear
107		Thanks a lot for your wishes on my birthday. Thanks you for making my birthday truly memorable.
108		Aight. I'll hit you up when I get some cash

3- Once done with labeling, save and export as a CSV

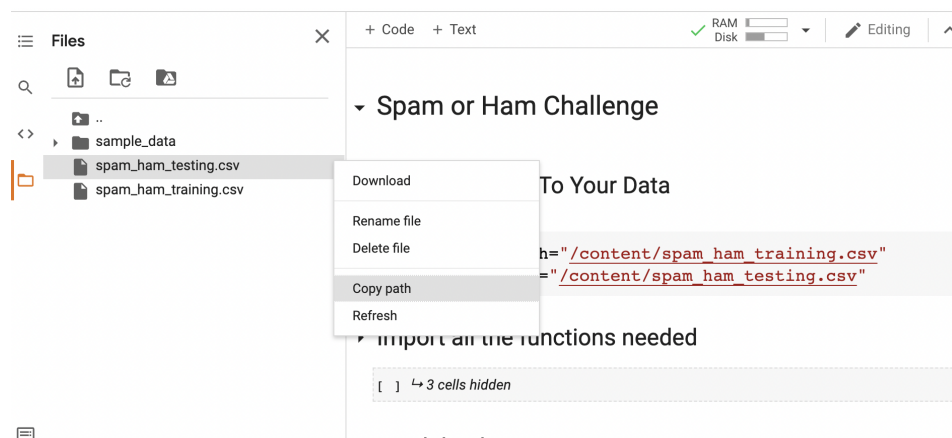
4- Open the [link](#) and make a copy of the notebook in your google drive



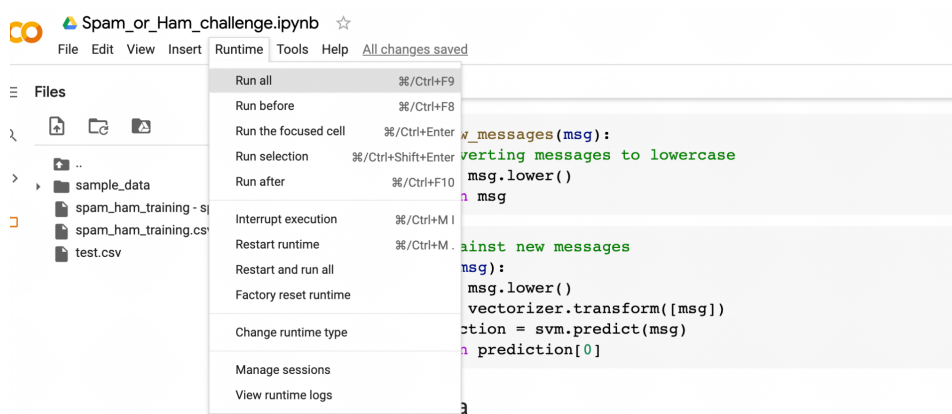
5- Once done, upload the test set CSV and your new training CSV to your colab environment as shown below



6- Once uploaded, copy the path of the training set to the variable call 'training_path' and the path of the test set to the variable called 'testing_path'. You can copy the menu by right-clicking on the csv (shown in the screenshot below)



7- Run all the notebook (as shown below) and submit the predictions, training set.



8- You can download the notebook as shown below:

CO

Spam_or_Ham_challenge.ipynb ☆

File Edit View Insert Runtime Tools Help All changes saved

File Explorer

Locate in Drive
Open in playground mode
New notebook
Open notebook ⌘/Ctrl+O
Upload notebook
Rename
Move
Move to trash
Save a copy in Drive
Save a copy as a GitHub Gist
Save a copy in GitHub
Save ⌘/Ctrl+S
Save and pin revision ⌘/Ctrl+M S
Revision history
Download
Print ⌘/Ctrl+P

Code + Text

```
33] def review_messages(msg):  
    # converting messages to lower  
    msg = msg.lower()  
    return msg  
  
34] # test against new messages  
def pred(msg):  
    msg = msg.lower()  
    msg = vectorizer.transform([msg])  
    prediction = svm.predict(msg)  
    return prediction[0]  
  
load the data  
  
35] data = pd.read_csv(training_path,  
                        delimiter=';',  
                        columns = ['v1', 'v2'])  
  
[36] ham_count=len(data[data['label']=='ham'])
```