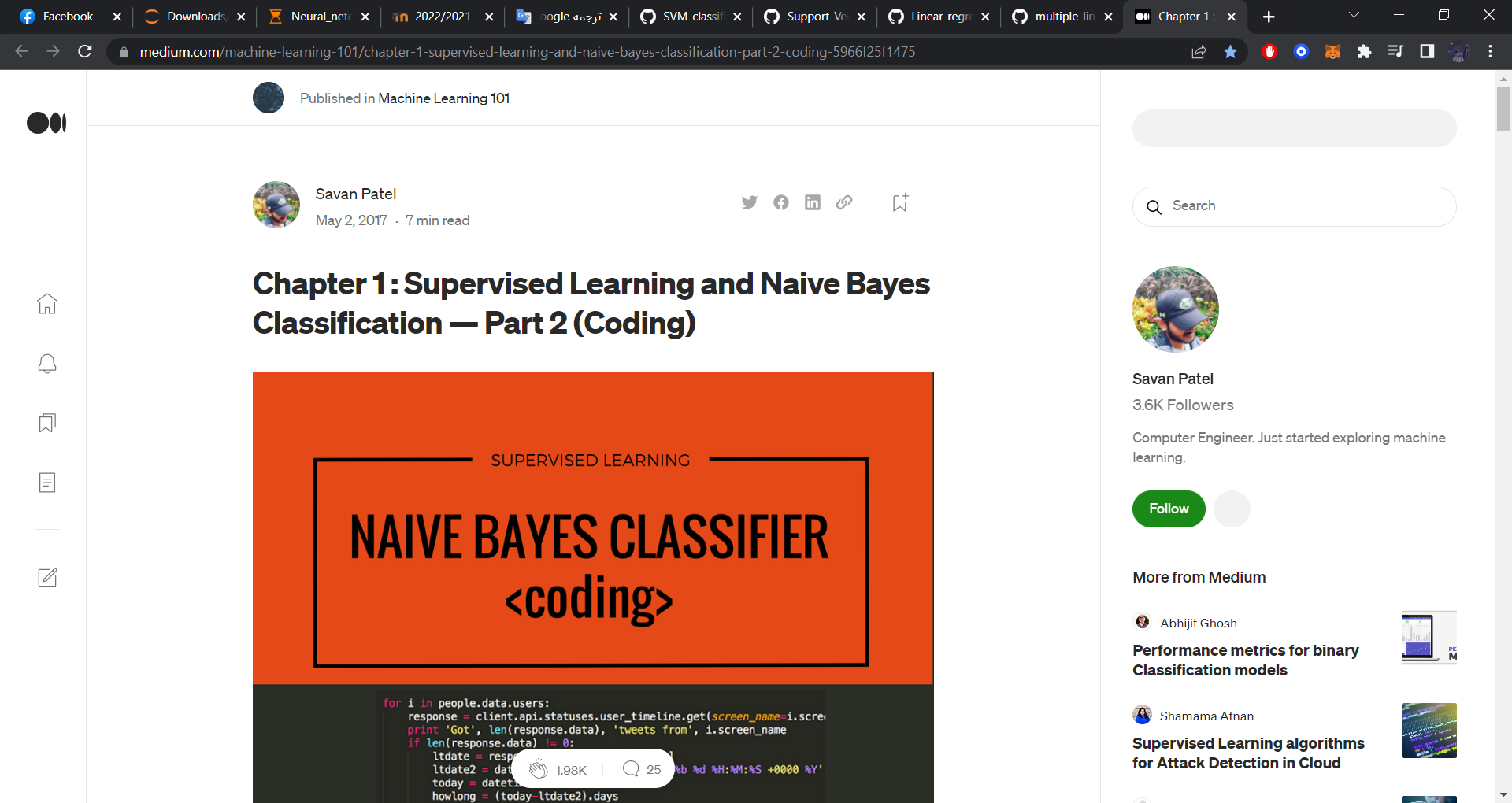
Naive Bayes (NB) classifier(PyCharm)

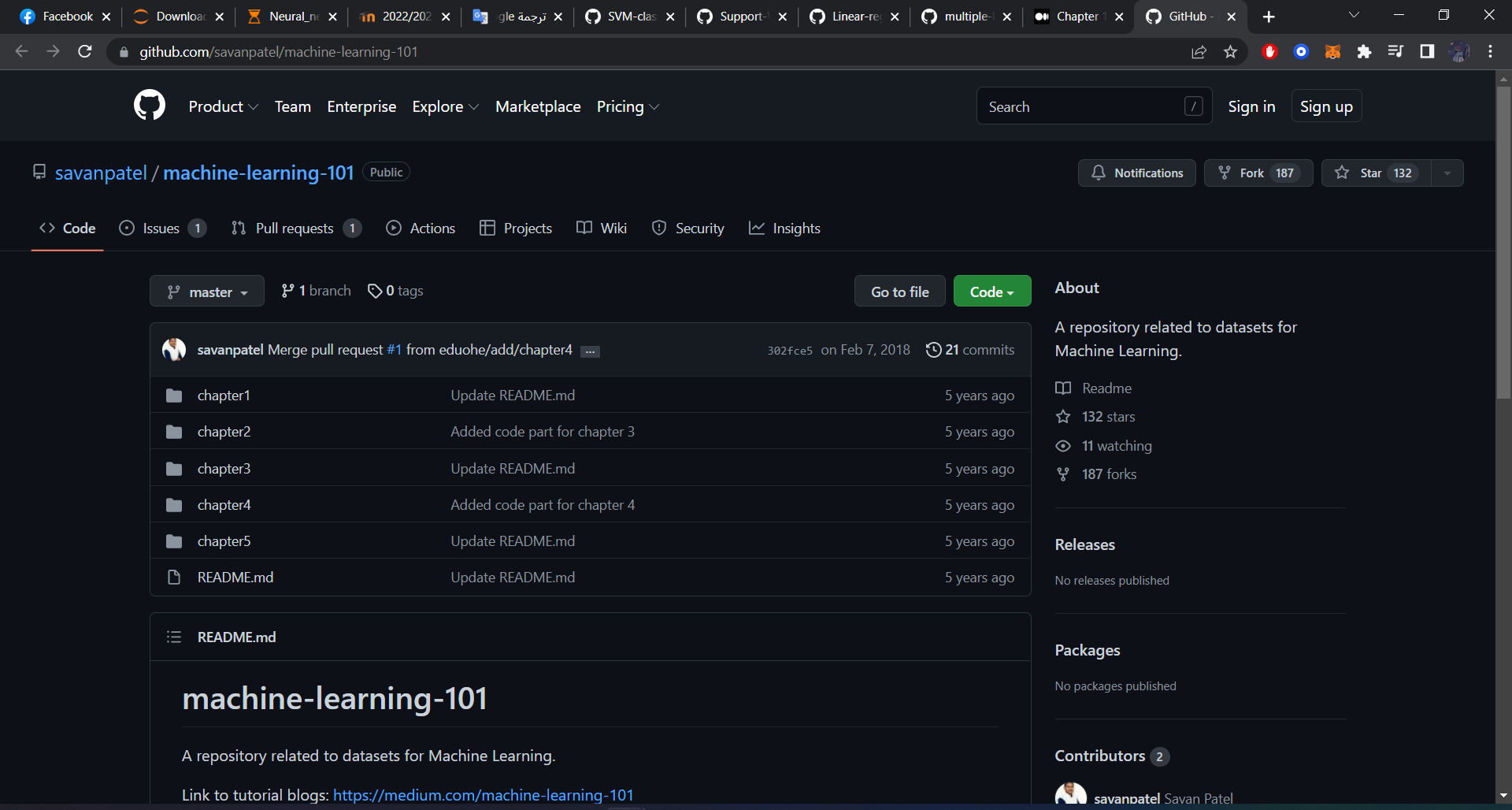
[1] I implement it from blog in medium site:

<https://medium.com/machine-learning-101/chapter-1-supervised-learning-and-naive-bayes-classification-part-2-coding-5966f25f1475>



[2] Download code and dataset from:

<https://github.com/savanpatel/machine-learning-101>



[3] About dataset : set of emails labelled as either from Spam or Not Spam.

[4] train set : 702 emails equally divided into spam and non spam category.

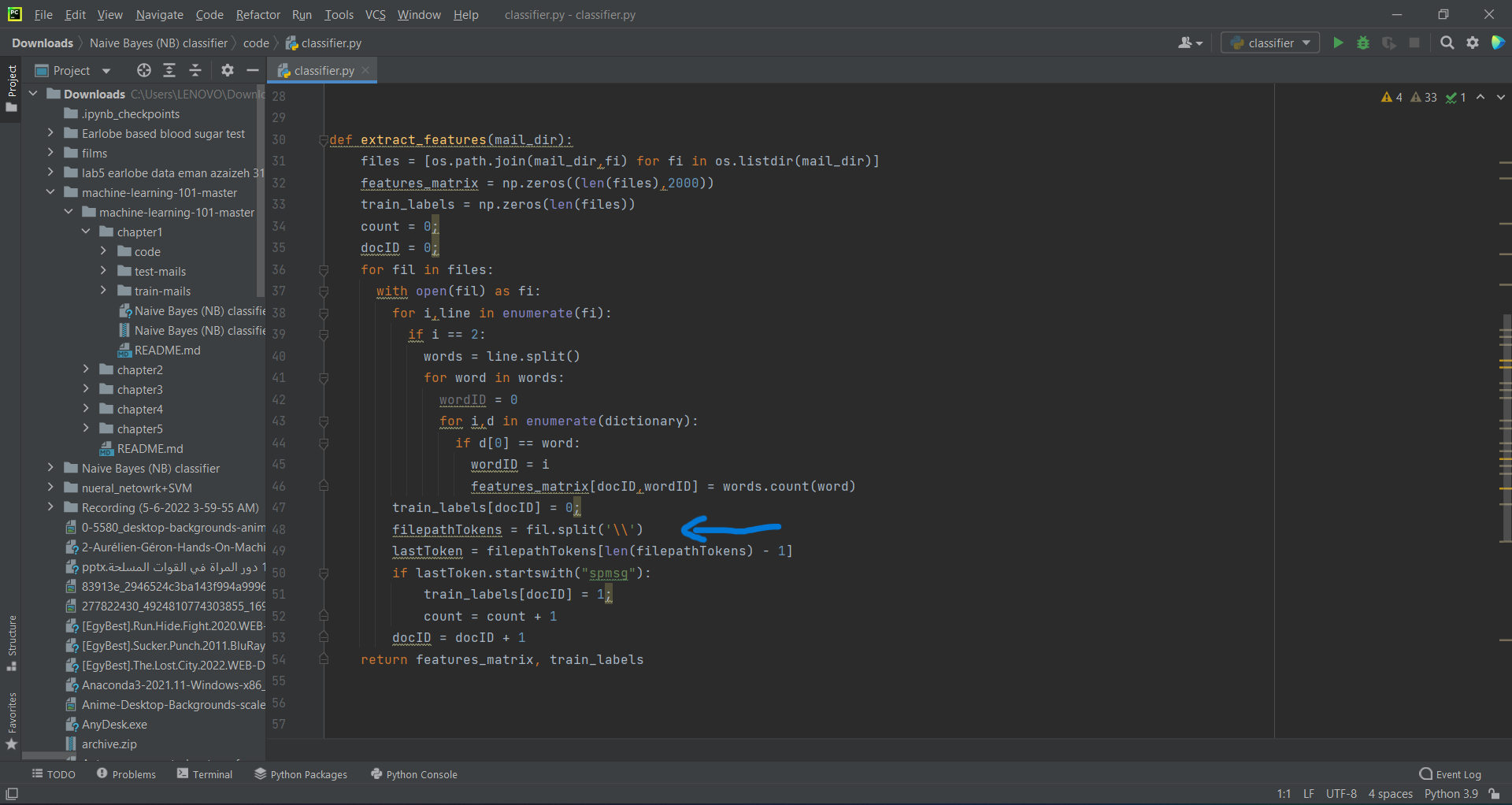
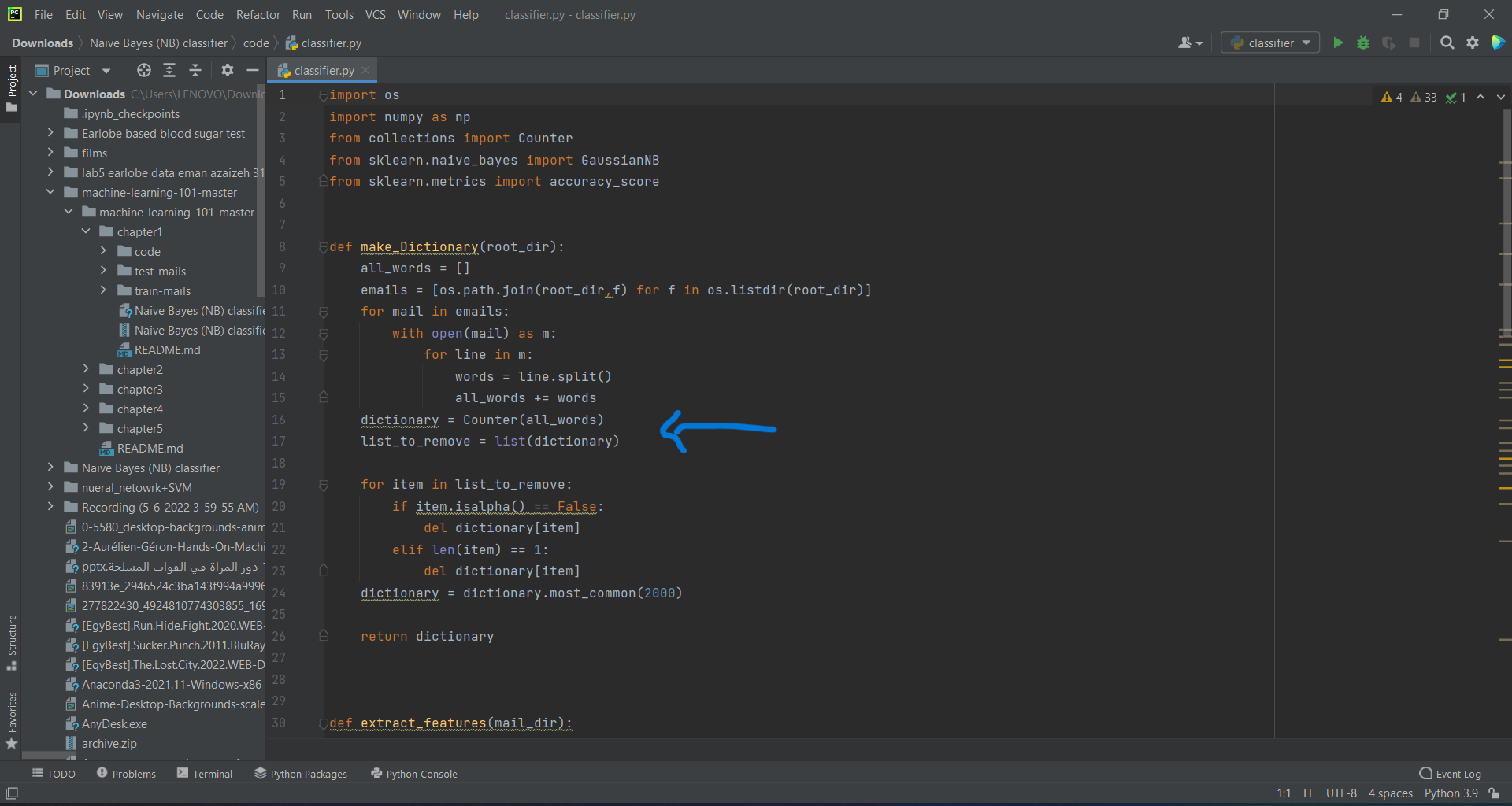
[5] test set : 260 emails.

[6] file names in two patterns:

number-numbermsg[number].txt : example 3-1msg1.txt (this are non spam emails).

OR

spmsg[Number].txt : example spmsga162.txt (these files are of spam emails).

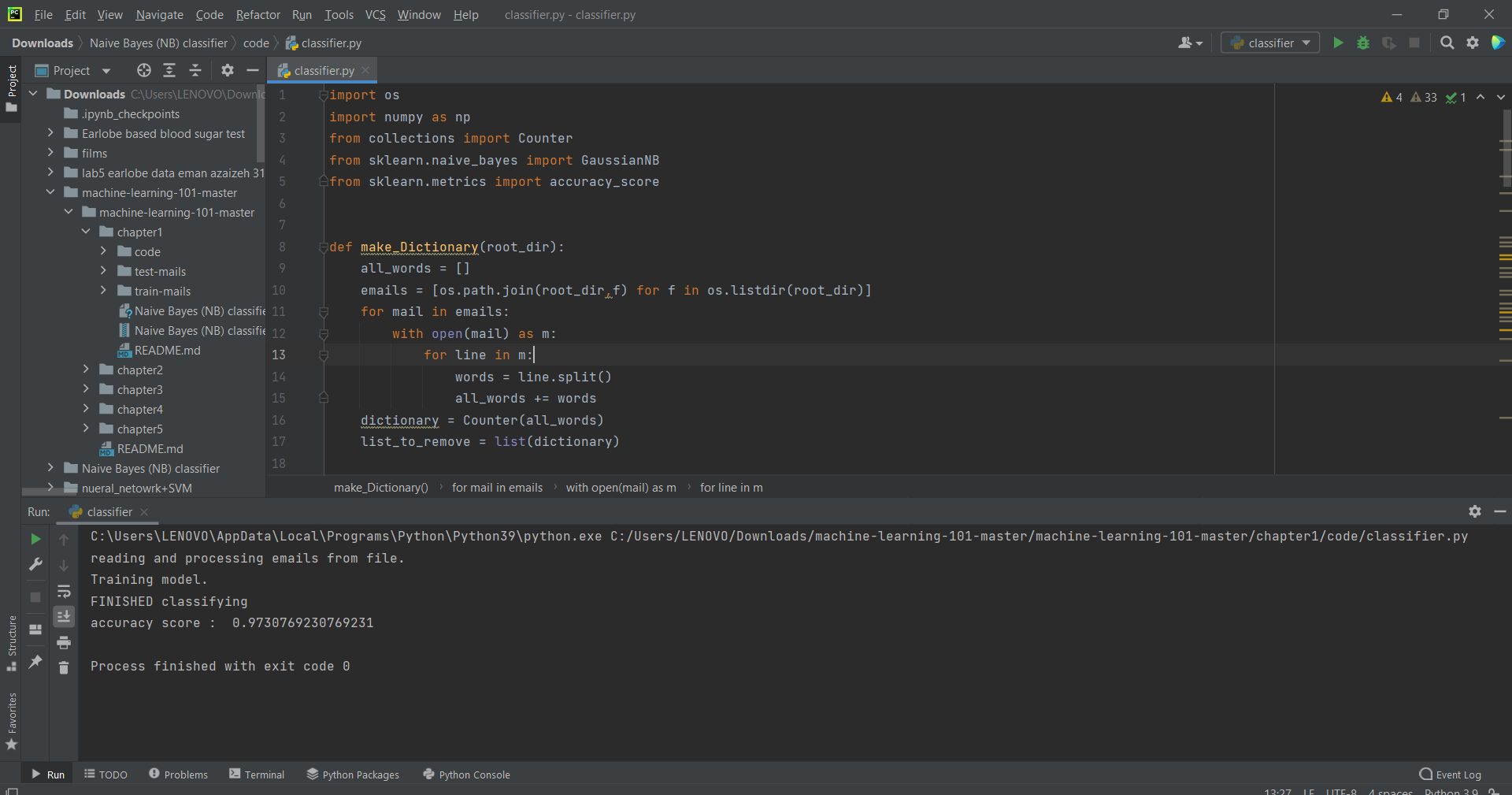
[7] replace this line of code [list\_to\_remove = dictionary.keys()] with [list\_to\_remove = list(dictionary)] if python version 3.x, and this line of code [filepathTokens = fil.split('/')] with [filepathTokens = fil.split('\\')] if running on W

[8] sklearn Naive Bayes provides three alternatives for model training: 1)Gaussian.

2)Multinomial.

3)Bernoulli.

[9]Run the code!



[9] train model with Gaussian and consider only most frequent 3000 words of dictionary from email,So I got this …

accuracy score : 0.961538461538.

- chaniging the most frequent words to 2000.

accuracy score : 0.9730769230769231.

- chaniging the most frequent words to 1500.

accuracy score : 0.961538461538.

-chaniging the most frequent words to 1000.

accuracy score : 0.8961538461538462.

- chaniging the most frequent words to 500.

accuracy score : 0.8153846153846154.

--------------------------------------------------------------------------

- chaniging the most frequent words to 3500.

accuracy score : 0.9615384615384616.

- chaniging the most frequent words to 4000.

accuracy score : 0.9576923076923077.

- chaniging the most frequent words to 4500.

accuracy score : 0.9576923076923077.

- chaniging the most frequent words to 5000.

accuracy score : 0.9538461538461539.

### Name : Ramah hashem madi

### ID no. : 31909303044