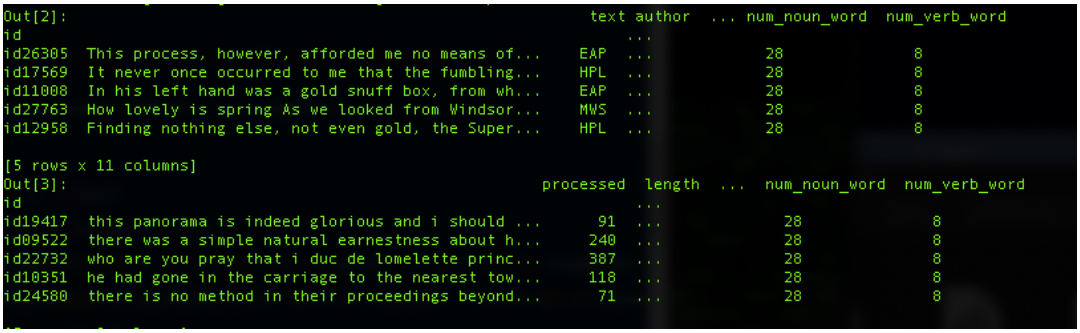
Kenneth Chuson
CSC 620-01 Natural Language Processing Spring 2022

HA#9 Report

a.) While adding a classifier at the end of the pipeline after the sklearn has been trained the text. The cross validation gives us enough flexibility and to determine how many tress and how deep should we go. After I changed from 5 to 3 cross validation I started to realize the hyper parameters has been changed little bit. And also the prediction changed a little bit, but it increase the mean prediction when I turned on to the 3 cross validation.

b.) 

```
Out[2]:                                                  text author  ... num_noun_word  num_verb_word
id                                                                    ...
id26305  This process, however, afforded me no means of...    EAP  ...           28              8
id17569  It never once occurred to me that the fumbling...    HPL  ...           28              8
id11008  In his left hand was a gold snuff box, from wh...    EAP  ...           28              8
id27763  How lovely is spring As we looked from Windsor...    MWS  ...           28              8
id12958  Finding nothing else, not even gold, the Super...    HPL  ...           28              8

[5 rows x 11 columns]
Out[3]:                                                processed  length  ...  num_noun_word  num_verb_word
id                                                                        ...
id19417  this panorama is indeed glorious and i should ...      91  ...           28              8
id09522  there was a simple natural earnestness about h...     240  ...           28              8
id22732  who are you pray that i duc de lomelette princ...     387  ...           28              8
id10351  he had gone in the carriage to the nearest tow...     118  ...           28              8
id24580  there is no method in their proceedings beyond...      71  ...           28              8
```

In the previous HA#8, I added number of adjectives, nouns, and verbs.

Class important and least important features I found.

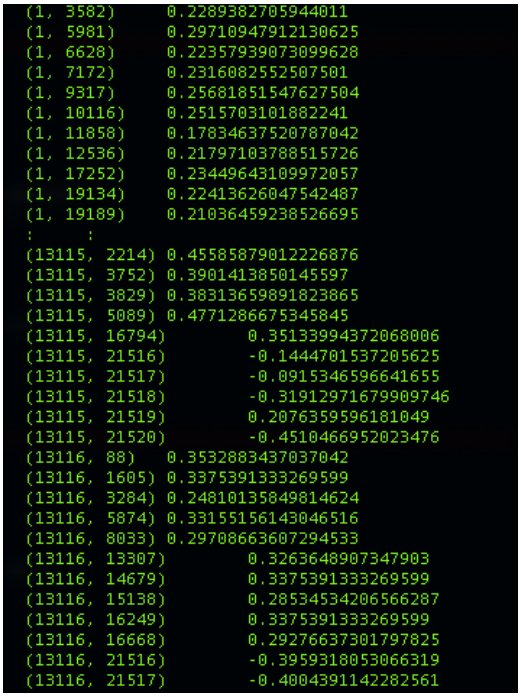
processed	length	words	words_not_stopword	avg_word_length
this process however afforded me no means of a...	224	41	21	6.380952
it never once occurred to me that the fumbling...	70	14	6	6.166667
in his left hand was a gold snuff box from whi...	195	36	19	5.947368
how lovely is				

Important features

commas	author
4	EAP
0	HPL
4	EAP

least important features

c.)



After training the data, some of the text have wrong prediction which if they have negative probabilities.