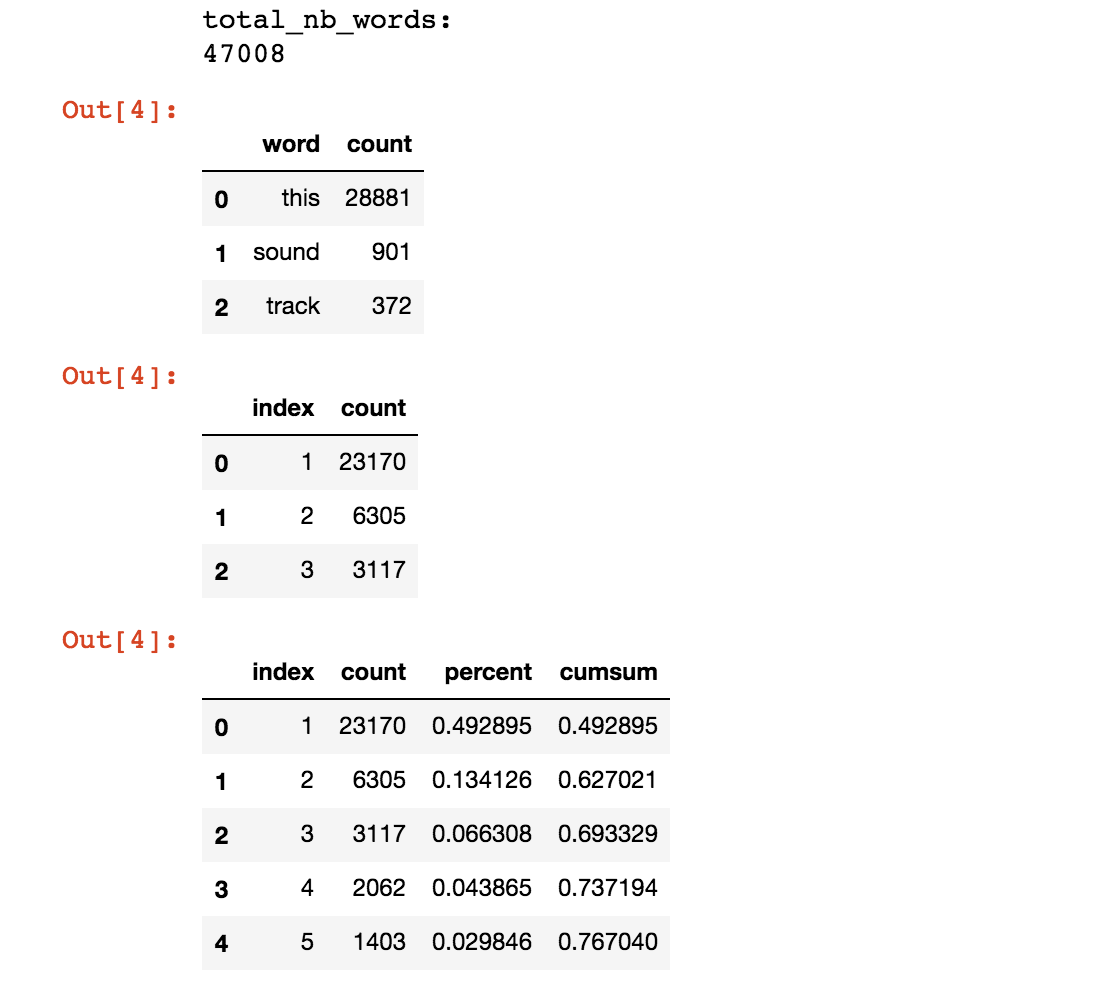
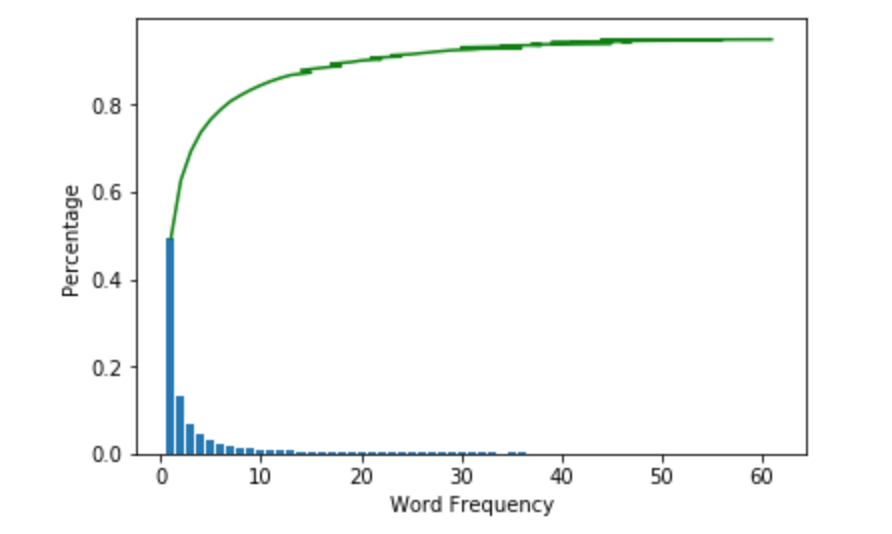
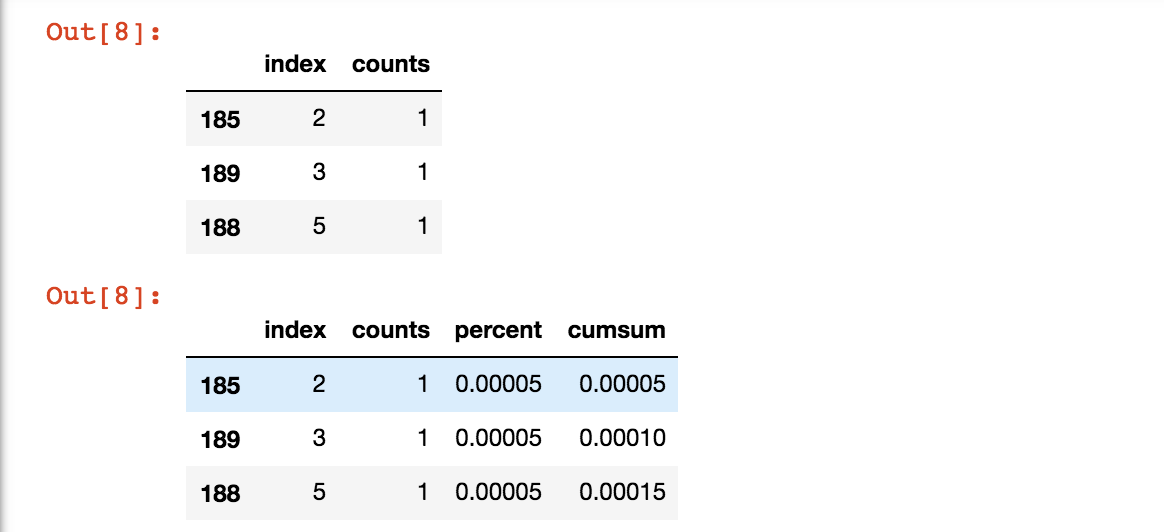
**Tune hyperparameters and observe how performance changes.**

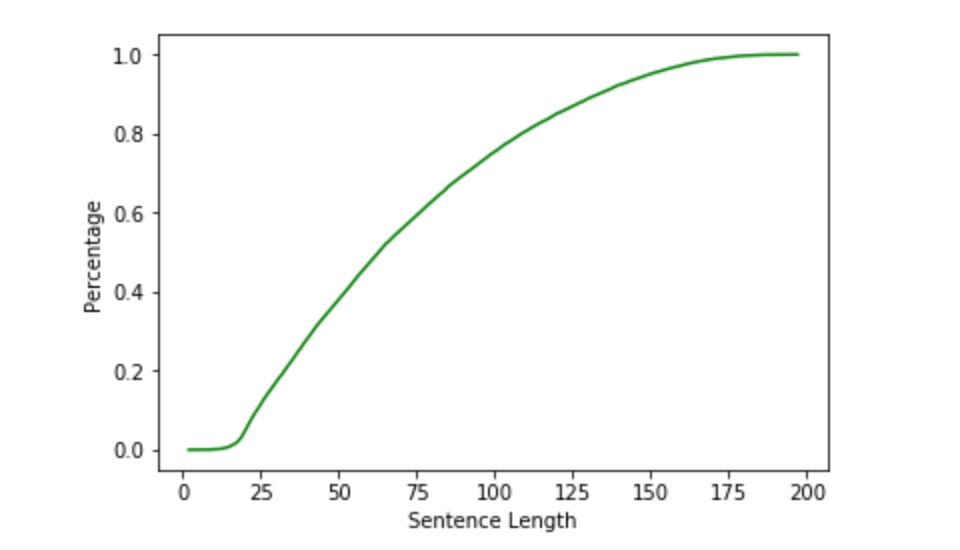
1. According to the chart below, you can find that >90% words appear in less than 30 times, which means if you like to include only words occur more than 30 times, then MAX\_NB\_WORDS = 10% \* total\_nb\_words = 4700. That is to say, if we set MAX\_NB\_WORDS > 4700, there is no obvious effect on the performance of precision. Instead, it only costs more time for computing on the word vectors. The reason is as follows: Firstly, it requires the same length of each vector. So it will need more time to pad the zero to the end of each vector. Secondly, it will also need more time to process the convolution layer since the vector is longer than before. Finally, there are more vectors to be process, so it costs longer time.





1. The chart below shows that 90% sentences have length<150, so it makes sense to set MAX\_DOC\_LEN=4~150. It only costs more time for computing if we set the MAX\_DOC\_LEN>150. The reason is as above.



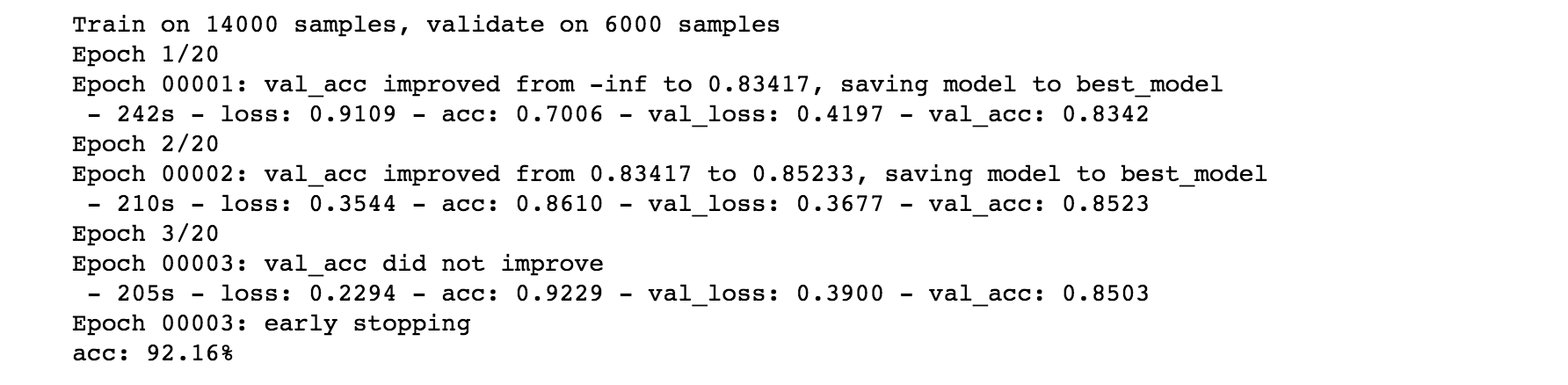


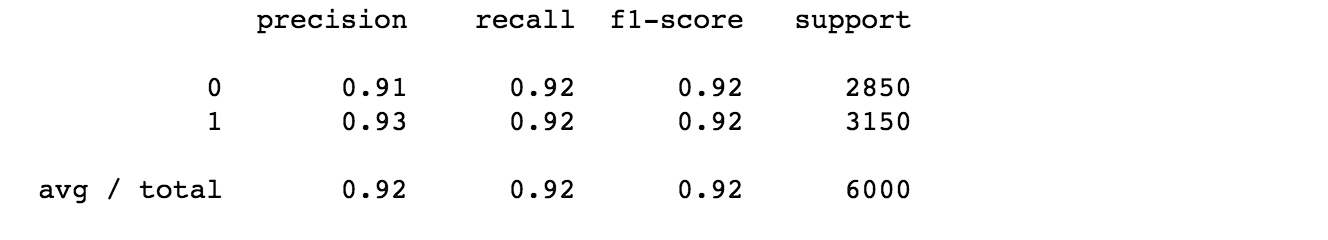
To prove the conclusion above, we give some example as follows:

**Case 1:**

MAX\_NB\_WORDS=10000

MAX\_DOC\_LEN=500

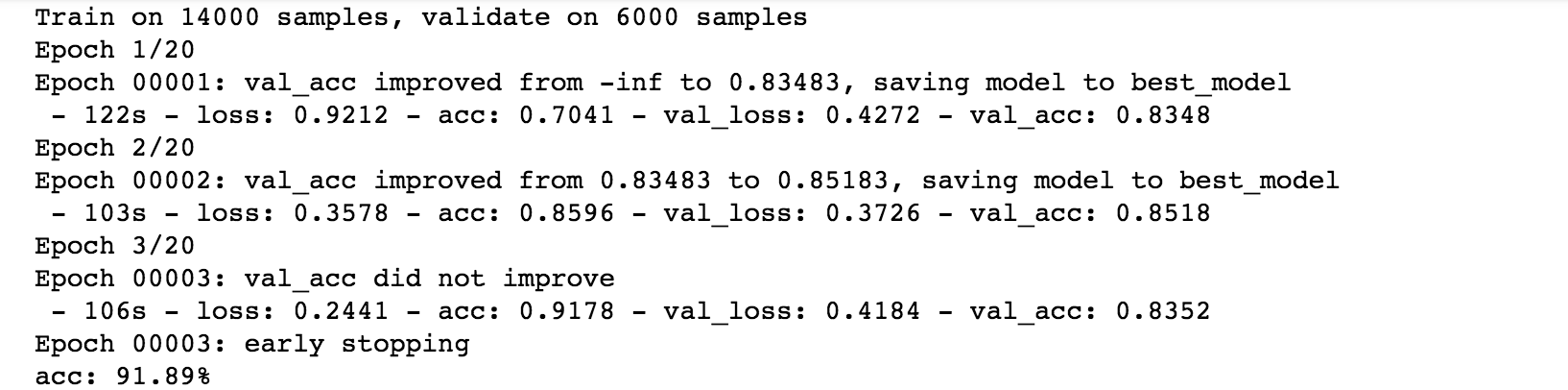


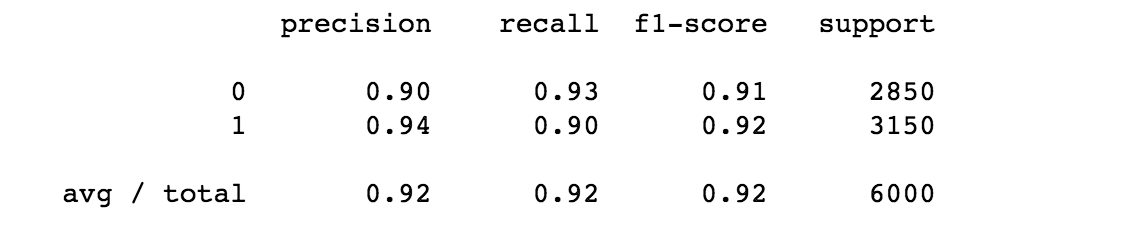


**Case 2:**

MAX\_NB\_WORDS=8000

MAX\_DOC\_LEN=200

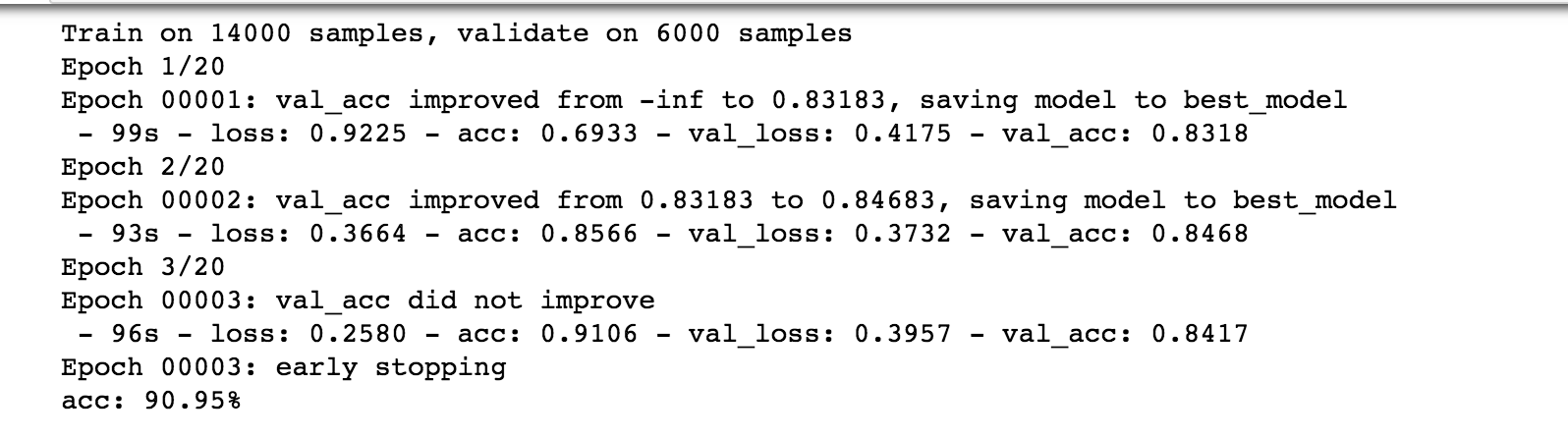


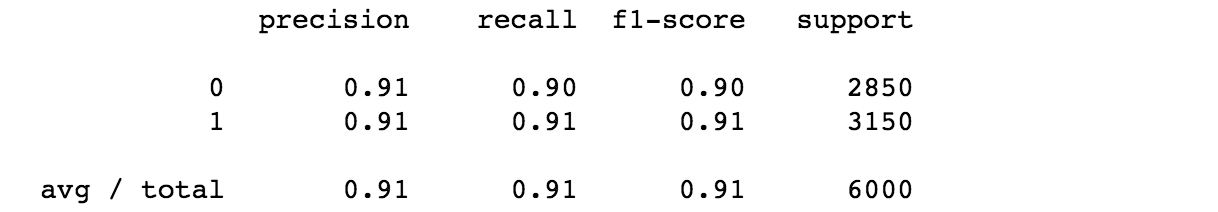


Case 3:

MAX\_NB\_WORDS=5000

MAX\_DOC\_LEN=200

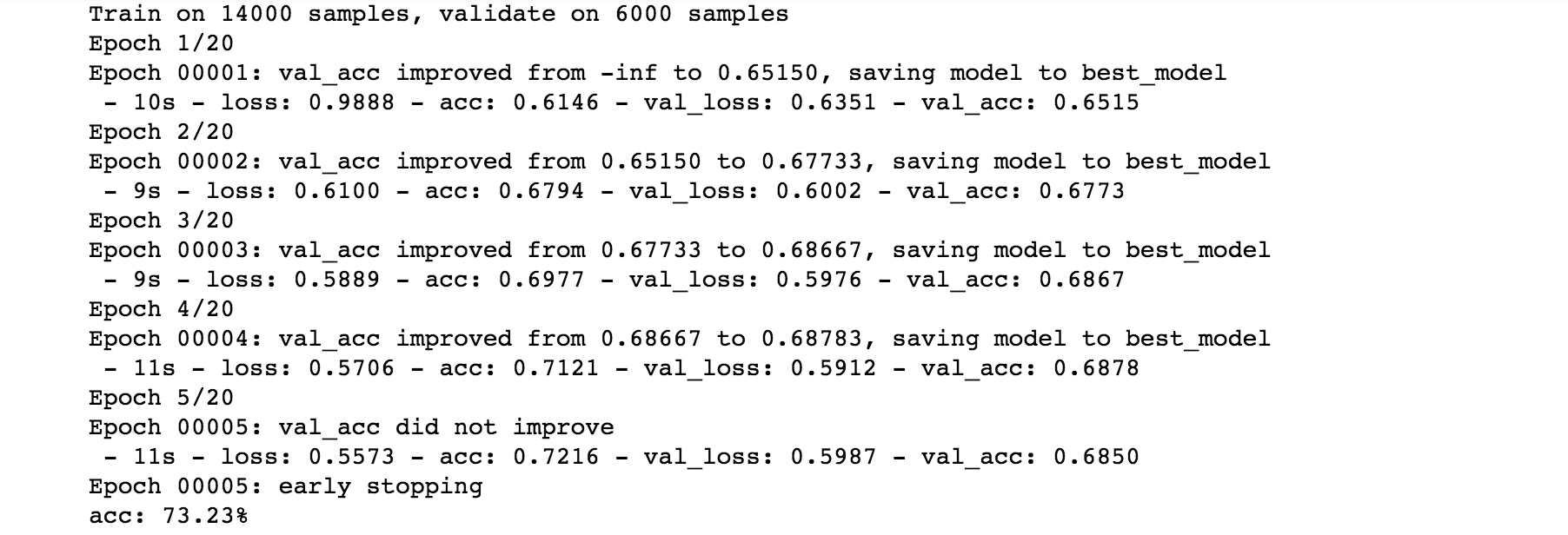


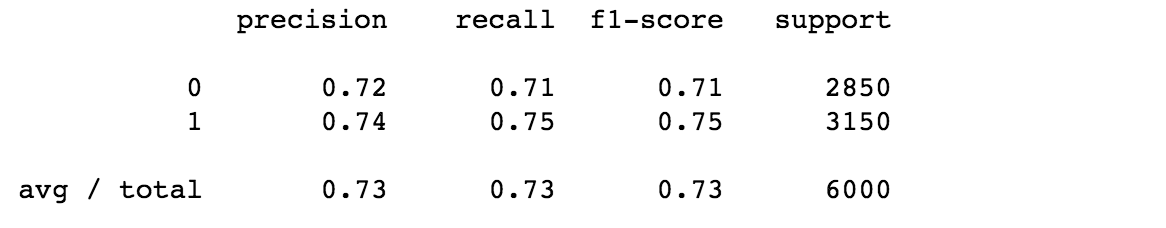


**Case 4:**

MAX\_NB\_WORDS=200

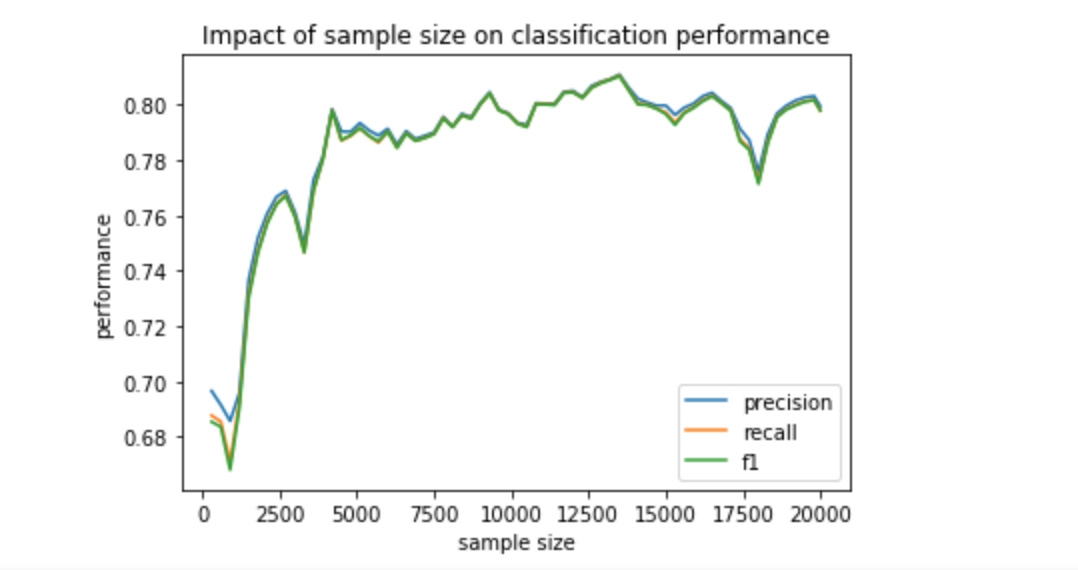
MAX\_DOC\_LEN=20





The data above shows, as the MAX\_NB\_WORDS and MAX\_DOC\_LEN getting larger, the running time for computing getting longer and the performance becoming higher. But while the two variables reach to a point, the performance stop getting higher while the running time continue growing.

**Compare the performance of CNN with the performance you obtained in Experiment 3 of Assignment 4.**



First of all, CNN is more precise and with better performance. The highest performance of CNN is higher than the highest performance of Naïve Bayes while the CNN is using the suitable hyper-parameters. And also the CNN reach to the highest performance using the less sample data. But again, the performance of CNN is not only base on the sample size, but also the hyper-parameters.