

Small but relevant clarification regarding my code and results:

As I mentioned in my report, in two of my Python scripts, the ones in which I am classifying over the mixed-genre set of 3000 sentences, I shuffle the set every time in the beginning of the code. Therefore, because of the way the code is written now, you will get different results every time you run the code.

However, this is not the case for the results I show in my report. They come from the same set of shuffled data, because of course I needed to compare results coming from the same observation.

Specifically, in “char_based.py”, I simply commented lines 13-17 after the first time I run the code, so that it would always consider the same “alldata_shuffled.txt”. Then I commented lines 14-18 in “wordbased.py” for the same reason.