









For full LM as a model as we can see that onfull test data the average mle score is 0.11

For offensive test data the average score is 0.08 and for not offensive test data it is 0.13 so here notffensive has shown greater likelihood.

For off LM as a model as we can see that onfull test data the average mle score is 0.23

For offensive test data the average score is 0.14 and for not offensive test data it is 0.26 so here notffensive has shown greater likelihood.

For not ff LM as a model as we can see that onfull test data the average mle score is 0.14

For offensive test data the average score is 0.11 and for not offensive test data it is 0.15 so here notffensive has shown greater likelihood.

Similarly, when we have took offensive LM it has created vocab mostly containing offensive words which is useful more in discriminating notoffensive data more and shown good score on full data as well.

When we are taking fullLM the maximum likelihood on offensive test data is less as compared to when we are taking notoffensive LM but the difference is not that much. As notoffensive LM would have created the not offensive vocab which might have been useful in discriminating the offensive data more accurately. It has shown greater average mle score for not offensive data as well.

I have created a train function in which preprocessed tweets are being passed list of tokens of each tweet is padded and bigrams are generated. The same function is used for creating LM model on full data, off data and not off data.

In the testing function I am dividing the test set into full, off and not off I am preprocessing the tweets and converting into bigrams and predicting the probability of bigrams by the fitted language model and multiplying the probabilities of each bigram in the specific tweet and appending it to each mletest1, mletest2 and mletest3.

I am ignoring the zero probabilities for now and I have not implemented smoothing in the probabilities calculation due to which the results might obviate from the enclosed above.