

CS671A Assignment 1

KSHITIZ SUMAN (14333)

(a) 1

- Every conversation ends with a punctuation like (,|.|?!).
- Every conversation starts with a newline or space before the single quote.
- Now by using `((. |\n)*?)` we can iterate through one by one character or newline and stop when any punctuation character and single quote is encountered.
- Finally, replaced the matched string by using the function `re.sub()` with the help of group capturing.

(a) 2

- Firstly, replaced the dot character of abbreviations like *Mr., Dr., G.K.C.* with a dummy character (*say@*).
- Every sentence ends with a punctuation character (?.|!)
- Inserted the tags using `re.sub()` module and then changed the dummy character inserted in the first step with the `'` character.

(b)

Using the output of the question (a) 2 , trained the algorithm. The model takes first 5 to 6 characters from left and right of each punctuation character , i.e.,left and right of each sentence termination . Stored the frequency of the character and by using SVM predicted the test data.The model has the accuracy is about 95.23%.