Assignment 2

Aim

Generate tweets: Use generated n-gram language model to generate tweets for a given entity (e.g. Mike Pence) or a topic (e.g., education) 1 . The length of the generated tweets are restricted to a number (say 10 words). Use the following method to write your algorithm:

(a) Random Generator: Randomly pick-up the n-grams that contains the topic or entity. Use that as pivot and add prefix and suffix to that.

Algorithm

1. Reverse the sentances in the corpus for the prefix generation with the original corpus file being used for suffix generation.

2. Initialise variables:

```
prefix_n = random.randint(1, numwords - 2)
suffix_n = numwords - prefix_n - 1
prefix = initialword
suffix = initialword
suffix_text = initialword
prefix_text = ''
```

3. Generate the prefix text prefix_n no. of times from the n gram:

```
for index in range(prefix_n):
        try:
                prefix = rev_cpd_2gram[prefix].generate()
                prefix_text = prefix + ' ' + prefix_text
        except Exception as e:
                print(e)
        return
```

4. Reverse prefix_text:

```
prefix_text = ' '.join(reversed(prefix_text.split()))
```

5. Generate suffix_text suffix_n no. of times from the n gram:

```
for index in range(suffix_n):
        try:
                suffix = cprob_2gram[suffix].generate()
                suffix_text = suffix_text + ' ' +suffix
        except Exception as e:
                print(e)
        return
```

6. Return sentance:

```
return prefix_text + ' ' + suffix_text
```

Top 5 tweets

input word: trump

1. will NAFTA and trump hate " —Hillary We must work

2. always was trump hate Mexicans We have drug prices and

3. we speculators wealthy the trump hate and for the American

4. will we sisters trump hate Join me in not incremental

5. will I someone needs Congress ISIL to trump hate —Hillary