

1 Spellcheck

January 18, 2024

1 Spellcheck

1.0.1 Import Libraries

```
[ ]: import string
import nltk

from nltk.corpus import brown, gutenberg
from nltk.corpus import stopwords
from nltk.util import ngrams
```

```
[ ]:
```

```

ques_data = ["Mr Patrick is our new principle","The company excepted all the
↳terms","Please don't keep your dog on the lose","The later is my best
↳friend","I need some stationary products for my craftwork","The actor
↳excepted the Oscar","I will call you later in the evening","Covid affects
↳the lungs","The council of the ministers were sworn in yesterday","Robert
↳too wants to accompany us to the park","Mia will council me about choosing
↳fashion as my career", "The bear at the zoo was very playful","The sheep
↳have a lot of fur that keeps them warm","The hot spring is at the furthest
↳corner of the street","Can you advice me on how to study for exams","The
↳team will loose the match if they don't play well","Can you go to the market
↳for me","The teachers asked the students to keep quite", "The heap of
↳garbage should be cleaned immediately","This is there house","Mr Patrick is
↳our new principal","The company accepted all the terms","Please don't keep
↳your dog on the loose","The latter is my best friend","I need some
↳stationery products for my craftwork","The actor accepted the Oscar","I will
↳call you later in the evening","Covid affects the lungs","The council of the
↳ministers were sworn in yesterday","Robert too wants to accompany us to the
↳park","Mia will counsel me about choosing fashion as my career", "The bear
↳at the zoo was very playful","The sheep have a lot of fur that keeps them
↳warm","The hot spring is at the farthest corner of the street","Can you
↳advise me on how to study for exams","The team will lose the match if they
↳don't play wel.", "Can you go to the market for me","The teachers asked the
↳students to keep quiet", "The heap of garbage should be cleaned
↳immediately","This is their house","Mr Patrick is our new principal","The
↳company accepted all the terms","Please don't keep your dog on the
↳loose","The latter is my best friend","I need some stationery products for
↳my craftwork","The actor accepted the Oscar","I will call you later in the
↳evening","Covid affects the lungs","The council of the ministers were sworn
↳in yesterday","Robert too wants to accompany us to the park","Mia will
↳counsel me about choosing fashion as my career", "The bear at the zoo was
↳very playful","The sheep have a lot of fur that keeps them warm","The hot
↳spring is at the farthest corner of the street","Can you advise me on how to
↳study for exams","The team will lose the match if they don't play well","Can
↳you go to the market for me","The teachers asked the students to keep
↳quiet", "The heap of garbage should be cleaned immediately","This is their
↳house"]

```

```

[ ]: ques_data.extend(ques_data)

ques_words = []
for i in ques_data:
    ques_words.extend(i.split())

```

1.0.2 Preprocessing the corpus

```
[ ]: corpus = brown.words() + gutenbergs.words() + ques_words
punc_stop = list(string.punctuation) + stopwords.words("english")

[ ]: lower = [x.lower() for x in corpus]
filter_words = [x for x in lower if x not in punc_stop]
filter_words = list(filter(lambda x: x.isalpha() and len(x) > 1, filter_words))
```

1.0.3 Building Bi-Gram Model

```
[ ]: bigrams = ngrams(filter_words, 2)
bi = list(bigrams)
bi[0:5]

[ ]: [('fulton', 'county'),
      ('county', 'grand'),
      ('grand', 'jury'),
      ('jury', 'said'),
      ('said', 'friday')]

[ ]: model = nltk.ConditionalFreqDist(bi)
```

1.0.4 Predictions

```
[ ]: def predict(question, model, choice1, choice2):
    processed = [i.lower() for i in question.split() if i not in punc_stop]

    for i in processed:
        if i in ["mask", "mask."]:
            break
    prev = i

    prob1 = model[prev].freq(choice1)
    prob2 = model[prev].freq(choice2)
    print(f"{choice1} = {prob1:.4f}\n{choice2} = {prob2:.4f}")

    if prob1 > prob2:
        return question.replace("MASK", choice1.upper())
    return question.replace("MASK", choice2.upper())
```

```
[ ]:
```

```

questions = ["Mr Patrick is our new MASK.", "The company MASK all the terms.
↳", "Please don't keep your dog on the MASK.", "The MASK is my best friend.",
↳ "I need some MASK products for my craftwork.", "The actor MASK the Oscar.",
↳ "I will call you MASK in the evening.", "Covid MASK the lungs.", "The MASK
↳ of the ministers were sworn in yesterday.", "Robert MASK wants to accompany
↳ us to the park.", "Mia will MASK me about choosing fashion as my career.",
↳ "The MASK at the zoo was very playful.", "The sheep have a lot of MASK that
↳ keeps them warm.", "The hot spring is at the MASK corner of the street.",
↳ "Can you MASK me on how to study for exams?", "The team will MASK the match
↳ if they don't play well.", "Can you go MASK the market for me?", "The
↳ teachers asked the students to keep MASK.", "The MASK of garbage should be
↳ cleaned immediately.", "This is MASK house."]

```

```

[ ]: choices = [["principle", "principal"], ["excepted", "accepted"], ["lose",
↳ "loose"], ["later", "latter"], ["stationary", "stationery"], ["excepted",
↳ "accepted"], ["later", "latter"], ["affects", "effects"], ["council",
↳ "counsel"], ["too", "to"], ["council", "counsel"], ["bare", "bear"], ["fur",
↳ "far"], ["furthest", "farthest"], ["advice", "advise"], ["loose", "lose"],
↳ ["to", "too"], ["quite", "quiet"], ["heap", "hip"], ["there", "their"]]

```

```

[ ]: for i in range(len(questions)):
    print(f"{i + 1}.", predict(questions[i], model, choices[i][0],
↳ choices[i][1]), "\n")

```

principle = 0.0012

principal = 0.0016

1. Mr Patrick is our new PRINCIPAL.

excepted = 0.0033

accepted = 0.0066

2. The company ACCEPTED all the terms.

lose = 0.0103

loose = 0.0205

3. Please don't keep your dog on the LOOSE.

later = 0.0000

latter = 0.0000

4. The LATTER is my best friend.

stationary = 0.0031

stationery = 0.0062

5. I need some STATIONERY products for my craftwork.

excepted = 0.0339

accepted = 0.0678

6. The actor ACCEPTED the Oscar.

later = 0.0071
 latter = 0.0000
 7. I will call you LATER in the evening.

affects = 1.0000
 effects = 0.0000
 8. Covid AFFECTS the lungs.

council = 0.0000
 counsel = 0.0000
 9. The COUNSEL of the ministers were sworn in yesterday.

too = 0.0000
 to = 0.0000
 10. Robert TO wants to accompany us to the park.

council = 0.3333
 counsel = 0.6667
 11. Mia will COUNSEL me about choosing fashion as my career.

bare = 0.0000
 bear = 0.0000
 12. The BEAR at the zoo was very playful.

fur = 0.0193
 far = 0.0000
 13. The sheep have a lot of FUR that keeps them warm.

furthest = 0.0064
 farthest = 0.0127
 14. The hot spring is at the FARTHEST corner of the street.

advice = 0.0000
 advise = 0.0000
 15. Can you ADVISE me on how to study for exams?

loose = 0.0222
 lose = 0.0444
 16. The team will LOSE the match if they don't play well.

to = 0.0000
 too = 0.0000
 17. Can you go TOO the market for me?

quite = 0.0030
 quiet = 0.0080
 18. The teachers asked the students to keep QUIET.

heap = 0.0000
hip = 0.0000
19. The HIP of garbage should be cleaned immediately.

there = 0.0000
their = 0.0000
20. This is THEIR house.