MINI PROJECT

THE PROBABILITY OF EACH WORD IN A TEXT

December 27, 2020

Anitta Susan Aniyan
Enrollment Number: EONFWL323567
Batch Number: 2020-4868

INPUT TEXT

The scientific study of probability is a modern development of mathematics. Gambling shows that there has been an interest in quantifying the ideas of probability for millennia, but exact mathematical descriptions arose much later. There are reasons for the slow development of the mathematics of probability. Whereas games of chance provided the impetus for the mathematical study of probability, fundamental issues are still obscured by the superstitions of gamblers.

According to Richard Jeffrey, "Before the middle of the seventeenth century, the term 'probable' (Latin probabilis) meant approvable, and was applied in that sense, univocally, to opinion and to action. A probable action or opinion was one such as sensible people would undertake or hold, in the circumstances." However, in legal contexts especially, 'probable' could also apply to propositions for which there was good evidence.

OBJECTIVES OF THE STUDY

1) To find every single unique word in a text.

2) To compute the set of neighbours for each word containing its next word, and

3) To calculate the probability of each word in the set.

CODES

The analysis is done by using Python software.

```
content =set()
unique_words = set()
content="""The scientific study of probability is a modern development of mathematics.
Gambling shows that there has been an interest in quantifying the ideas of probability for
millennia, but exact mathematical descriptions arose much later. There are reasons for the
slow development of the mathematics of probability. Whereas games of chance provided the
impetus for the mathematical study of probability, fundamental issues are still obscured by
the superstitions of gamblers.
According to Richard Jeffrey, "Before the middle of the seventeenth century, the term 'proba-
ble' (Latin probabilis) meant approvable, and was applied in that sense, univocally, to opinion
and to action. A probable action or opinion was one such as sensible people would undertake
or hold, in the circumstances." However, in legal contexts especially, 'probable' could also
apply to propositions for which there was good evidence."""
```

```
    for char in ',. "()    n':
content=content.replace(char,' ')
content=content.lower().split()
d ={}
for index, word in enumerate(content):
if index == len(content) - 1:
d[word] = d.get(word, [])
continue
res = d.get(word, [])
res.append(content[index+1])
d[word] = res
print(k: len(v) for k, v in d.items())
print(k: v for k, v in d.items())
for k, v in d.items():
print(" of neighbourhood of -", "'"+k+"'", end=" n  t")
if len(v)==0:
print("Not Applicable")
for rep in list(dict.from keys(v)):
print("Probability of ", "'"+rep+"'",":",
v.count(rep)/len(v), end=" n   t")
```

OUTPUT

1. *For set of unique words with count of each word in a dictionary*:

'the': 11, 'scientific': 1, 'study': 2, 'of': 9, 'probability': 4, 'is': 1, 'a': 2, 'modern': 1, 'development': 2, 'mathematics': 2, 'gambling': 1, 'shows': 1, 'that': 2, 'there': 3, 'has': 1, 'been': 1, 'an': 1, 'interest': 1, 'in': 4, 'quantifying': 1, 'ideas': 1, 'for': 4, 'millennia': 1, 'but': 1, 'exact': 1, 'mathematical': 2, 'descriptions': 1, 'arose': 1, 'much': 1, 'later': 1, 'are': 2, 'reasons': 1, 'slow': 1, 'whereas': 1, 'games': 1, 'chance': 1, 'provided': 1, 'impetus': 1, 'fundamental': 1, 'issues': 1, 'still': 1, 'obscured': 1, 'by': 1, 'superstitions': 1, 'gamblers': 1, 'according': 1, 'to': 4, 'richard': 1, 'jeffrey': 1, 'before': 1, 'middle': 1, 'seventeenth': 1, 'century': 1, 'term': 1, 'probable': 3, 'latinâ': 1, 'probabilis': 1, 'meantâ': 1, 'approvable': 1, 'and': 2, 'was': 3, 'applied': 1, 'sense': 1, 'univocally': 1, 'opinion': 2, 'action': 2, 'or': 2, 'one': 1, 'such': 1, 'as': 1, 'sensible': 1, 'people': 1, 'would': 1, 'undertake': 1, 'hold': 1, 'circumstances': 1, 'however': 1, 'legal': 1, 'contexts': 1,

'especially': 1, 'could': 1, 'also': 1, 'apply': 1, 'propositions': 1, 'which': 1, 'good': 1, 'evidence':1

2. *Computation of the set of neighbours for each word containing its next wordagain in*

'the': ['scientific', 'ideas', 'slow', 'mathematics', 'impetus', 'mathematical', 'superstitions', 'middle', 'seventeenth', 'term', 'circumstances'],

'scientific': ['study'],

'study': ['of', 'of'],

'of': ['probability', 'mathematics', 'probability', 'the', 'probability', 'chance', 'probability', 'gamblers', 'the'],

'probability': ['is', 'for', 'whereas', 'fundamental'],

'is': ['a'],

'a': ['modern', 'probable'],

'modern': ['development'],

'development': ['of', 'of'],

'mathematics': ['gambling', 'of'],

'gambling': ['shows'],

'shows': ['that'],

'that': ['there', 'sense'],

'there': ['has', 'are', 'was'],

'has': ['been'],

'been': ['an'],

'an': ['interest'],

'interest': ['in'],

'in': ['quantifying', 'that', 'the', 'legal'],

'quantifying': ['the'],

'ideas': ['of'],

'for': ['millennia', 'the', 'the', 'which'],

'millennia': ['but'],

'but': ['exact'],

'exact': ['mathematical'],

'mathematical': ['descriptions', 'study'],

'descriptions': ['arose'],

'arose': ['much'],

'much': ['later'],

'later': ['there'],

'are': ['reasons', 'still'],

'reasons': ['for'],

'slow': ['development'],

'whereas': ['games'],

'games': ['of'],

'chance': ['provided'],

'provided': ['the'],

'impetus': ['for'],

'fundamental': ['issues'],

'issues': ['are'],

'still': ['obscured'],

'obscured': ['by'],

'by': ['the'],

'superstitions': ['of'],

'gamblers': ['according'],

'according': ['to'],

'to': ['richard', 'opinion', 'action', 'propositions'],

'richard': ['jeffrey'],

'jeffrey': ['before'],

'before': ['the'],

'middle': ['of'],

'seventeenth': ['century'],

'century': ['the'],

'term': ['probable'],

'probable': ['latinâ', 'action', 'could'],

'latinâ': ['probabilis'],

'probabilis': ['meantâ'],

'meantâ': ['approvable'],

'approvable': ['and'],

'and': ['was', 'to'],

'was': ['applied', 'one', 'good'],

'applied': ['in'],

'sense': ['univocally'],

'univocally': ['to'],

'opinion': ['and', 'was'],

'action': ['a', 'or'],

```
'or': ['opinion', 'hold'],

'one': ['such'],

'such': ['as'],

'as': ['sensible'],

'sensible': ['people'],

'people': ['would'],

'would': ['undertake'],

'undertake': ['or'],

'hold': ['in'],

'circumstances': ['however'],

'however': ['in'],

'legal': ['contexts'],

'contexts': ['especially'],

'especially': ['probable'],

'could': ['also'],

'also': ['apply'],

'apply': ['to'],

'propositions': ['for'],
```

'which': ['there'],

'good': ['evidence'],

'evidence': []

3. *The probability of each word in the each set*:

Set of neighbourhood of - 'the'
Probability of 'scientific' : 0.09090909090909091
Probability of 'ideas' : 0.09090909090909091
Probability of 'slow' : 0.09090909090909091
Probability of 'mathematics' : 0.09090909090909091
Probability of 'impetus' : 0.09090909090909091
Probability of 'mathematical' : 0.09090909090909091
Probability of 'superstitions' : 0.09090909090909091
Probability of 'middle' : 0.09090909090909091
Probability of 'seventeenth' : 0.09090909090909091
Probability of 'term' : 0.09090909090909091
Probability of 'circumstances' : 0.09090909090909091

Set of neighbourhood of - 'scientific'
Probability of 'study' : 1.0

Set of neighbourhood of - 'study'
Probability of 'of' : 1.0

Set of neighbourhood of - 'of'
Probability of 'probability' : 0.4444444444444444
Probability of 'mathematics' : 0.1111111111111111
Probability of 'the' : 0.2222222222222222
Probability of 'chance' : 0.1111111111111111
Probability of 'gamblers' : 0.1111111111111111

Set of neighbourhood of - 'probability'
Probability of 'is' : 0.25
Probability of 'for' : 0.25
Probability of 'whereas' : 0.25
Probability of 'fundamental' : 0.25

Set of neighbourhood of - 'is'
Probability of 'a' : 1.0

Set of neighbourhood of - 'a'
Probability of 'modern' : 0.5
Probability of 'probable' : 0.5

Set of neighbourhood of - 'modern'
Probability of 'development' : 1.0

Set of neighbourhood of - 'development'
Probability of 'of' : 1.0

Set of neighbourhood of - 'mathematics'
Probability of 'gambling' : 0.5
Probability of 'of' : 0.5

Set of neighbourhood of - 'gambling'
Probability of 'shows' : 1.0

Set of neighbourhood of - 'shows'
Probability of 'that' : 1.0

Set of neighbourhood of - 'that'
Probability of 'there' : 0.5
Probability of 'sense' : 0.5

Set of neighbourhood of - 'there'
Probability of 'has' : 0.3333333333333333

Probability of 'are' : 0.3333333333333333
Probability of 'was' : 0.3333333333333333

Set of neighbourhood of - 'has'
Probability of 'been' : 1.0

Set of neighbourhood of - 'been'
Probability of 'an' : 1.0

Set of neighbourhood of - 'an'
Probability of 'interest' : 1.0

Set of neighbourhood of - 'interest'
Probability of 'in' : 1.0

Set of neighbourhood of - 'in'
Probability of 'quantifying' : 0.25
Probability of 'that' : 0.25
Probability of 'the' : 0.25
Probability of 'legal' : 0.25

Set of neighbourhood of - 'quantifying'
Probability of 'the' : 1.0

Set of neighbourhood of - 'ideas'
Probability of 'of' : 1.0

Set of neighbourhood of - 'for'
Probability of 'millennia' : 0.25
Probability of 'the' : 0.5
Probability of 'which' : 0.25

Set of neighbourhood of - 'millennia'
Probability of 'but' : 1.0

Set of neighbourhood of - 'but'
Probability of 'exact' : 1.0

Set of neighbourhood of - 'exact'
Probability of 'mathematical' : 1.0

Set of neighbourhood of - 'mathematical'
Probability of 'descriptions' : 0.5
Probability of 'study' : 0.5

Set of neighbourhood of - 'descriptions'
Probability of 'arose' : 1.0

Set of neighbourhood of - 'arose'
Probability of 'much' : 1.0

Set of neighbourhood of - 'much'
Probability of 'later' : 1.0

Set of neighbourhood of - 'later'
Probability of 'there' : 1.0

Set of neighbourhood of - 'are'
Probability of 'reasons' : 0.5
Probability of 'still' : 0.5

Set of neighbourhood of - 'reasons'
Probability of 'for' : 1.0

Set of neighbourhood of - 'slow'
Probability of 'development' : 1.0

Set of neighbourhood of - 'whereas'
Probability of 'games' : 1.0

Set of neighbourhood of - 'games'
Probability of 'of' : 1.0

Set of neighbourhood of - 'chance'
Probability of 'provided' : 1.0

Set of neighbourhood of - 'provided'
Probability of 'the' : 1.0

Set of neighbourhood of - 'impetus'
Probability of 'for' : 1.0

Set of neighbourhood of - 'fundamental'
Probability of 'issues' : 1.0

Set of neighbourhood of - 'issues'
Probability of 'are' : 1.0

Set of neighbourhood of - 'still'
Probability of 'obscured' : 1.0

Set of neighbourhood of - 'obscured'
Probability of 'by' : 1.0

Set of neighbourhood of - 'by'
Probability of 'the' : 1.0

Set of neighbourhood of - 'superstitions'
Probability of 'of' : 1.0

Set of neighbourhood of - 'gamblers'
Probability of 'according' : 1.0

Set of neighbourhood of - 'according'
Probability of 'to' : 1.0

Set of neighbourhood of - 'to'
Probability of 'richard' : 0.25
Probability of 'opinion' : 0.25
Probability of 'action' : 0.25
Probability of 'propositions' : 0.25

Set of neighbourhood of - 'richard'
Probability of 'jeffrey' : 1.0

Set of neighbourhood of - 'jeffrey'
Probability of 'before' : 1.0

Set of neighbourhood of - 'before'
Probability of 'the' : 1.0

Set of neighbourhood of - 'middle'
Probability of 'of' : 1.0

Set of neighbourhood of - 'seventeenth'
Probability of 'century' : 1.0

Set of neighbourhood of - 'century'
Probability of 'the' : 1.0

Set of neighbourhood of - 'term'
Probability of 'probable' : 1.0

Set of neighbourhood of - 'probable'
Probability of 'latinâ' : 0.3333333333333333
Probability of 'action' : 0.3333333333333333
Probability of 'could' : 0.3333333333333333

Set of neighbourhood of - 'latinâ'
Probability of 'probabilis' : 1.0

Set of neighbourhood of - 'probabilis'
Probability of 'meantâ' : 1.0

Set of neighbourhood of - 'meantâ'
Probability of 'approvable' : 1.0

Set of neighbourhood of - 'approvable'
Probability of 'and' : 1.0

Set of neighbourhood of - 'and'
Probability of 'was' : 0.5
Probability of 'to' : 0.5

Set of neighbourhood of - 'was'
Probability of 'applied' : 0.3333333333333333
Probability of 'one' : 0.3333333333333333
Probability of 'good' : 0.3333333333333333

Set of neighbourhood of - 'applied'
Probability of 'in' : 1.0

Set of neighbourhood of - 'sense'
Probability of 'univocally' : 1.0

Set of neighbourhood of - 'univocally'
Probability of 'to' : 1.0

Set of neighbourhood of - 'opinion'
Probability of 'and' : 0.5
Probability of 'was' : 0.5

Set of neighbourhood of - 'action'
Probability of 'a' : 0.5
Probability of 'or' : 0.5

Set of neighbourhood of - 'or'
Probability of 'opinion' : 0.5
Probability of 'hold' : 0.5

Set of neighbourhood of - 'one'
Probability of 'such' : 1.0

Set of neighbourhood of - 'such'
Probability of 'as' : 1.0

Set of neighbourhood of - 'as'
Probability of 'sensible' : 1.0

Set of neighbourhood of - 'sensible'
Probability of 'people' : 1.0

Set of neighbourhood of - 'people'
Probability of 'would' : 1.0

Set of neighbourhood of - 'would'
Probability of 'undertake' : 1.0
Set of neighbourhood of - 'undertake'
Probability of 'or' : 1.0

Set of neighbourhood of - 'hold'
Probability of 'in' : 1.0

Set of neighbourhood of - 'circumstances'
Probability of 'however' : 1.0

Set of neighbourhood of - 'however'
Probability of 'in' : 1.0

Set of neighbourhood of - 'legal'

Probability of 'contexts' : 1.0

Set of neighbourhood of - 'contexts'
Probability of 'especially' : 1.0

Set of neighbourhood of - 'especially'
Probability of 'probable' : 1.0

Set of neighbourhood of - 'could'
Probability of 'also' : 1.0

Set of neighbourhood of - 'also'
Probability of 'apply' : 1.0

Set of neighbourhood of - 'apply'
Probability of 'to' : 1.0

Set of neighbourhood of - 'propositions'
Probability of 'for' : 1.0

Set of neighbourhood of - 'which'
Probability of 'there' : 1.0

Set of neighbourhood of - 'good'
Probability of 'evidence' : 1.0

Set of neighbourhood of - 'evidence'
Not Applicable