

CSE6060

Statistical Natural Language Processing

Hands on Task

Name : Kavianand G

Reg. No. : 19MAI0050

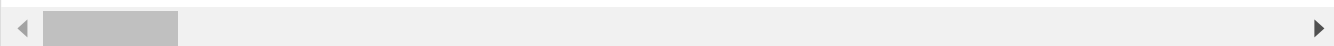
Date : 19 – June – 2020

Bigrams & Collocations

Dataset - Gettysburg address

In [1]:

```
1 data="""Four score and seven years ago our fathers brought forth on this continent, a r
2
3 Now we are engaged in a great civil war, testing whether that nation, or any nation so
4 We have come to dedicate a portion of that field, as a final resting place for those wh
5 It is altogether fitting and proper that we should do this. But, in a larger sense, we
6 The brave men, living and dead, who struggled here, have consecrated it, far above our
```



In [2]:

```
1 data
```

Out[2]:

```
'Four score and seven years ago our fathers brought forth on this continent,
a new nation, conceived in Liberty, and dedicated to the proposition that al
l men are created equal.\n\nNow we are engaged in a great civil war, testing
whether that nation, or any nation so conceived and so dedicated, can long e
ndure. We are met on a great battle-field of that war. \nWe have come to ded
icate a portion of that field, as a final resting place for those who here g
ave their lives that that nation might live. \nIt is altogether fitting and
proper that we should do this. But, in a larger sense, we can not dedicate -
- we can not consecrate -- we can not hallow -- this ground. The brave men,
living and dead, who struggled here, have consecrated it, far above our poor
power to add or detract. The world will little note, nor long remember what
we say here, but it can never forget what they did here. It is for us the li
ving, rather, to be dedicated here to the unfinished work which they who fou
ght here have thus far so nobly advanced. It is rather for us to be here ded
icated to the great task remaining before us -- that from these honored dead
we take increased devotion to that cause for which they gave the last full m
easure of devotion -- that we here highly resolve that these dead shall not
have died in vain -- that this nation, under God, shall have a new birth of
freedom -- and that government of the people, by the people, for the people,
shall not perish from the earth.'
```

Remove Stopwords from the text

In [3]:

```
1 # Datasetimport nltk
2 from nltk.corpus import stopwords
3 from nltk.tokenize import word_tokenize
4
5 stop_words = (stopwords.words('english'))
6 word_tokens = word_tokenize(data)
```

In [4]:

```
1 print(stop_words)
```

```
['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', "you'r  
e", "you've", "you'll", "you'd", 'your', 'yours', 'yourself', 'yourselves',  
'he', 'him', 'his', 'himself', 'she', "she's", 'her', 'hers', 'herself', 'i  
t', "it's", 'its', 'itself', 'they', 'them', 'their', 'theirs', 'themselv  
s', 'what', 'which', 'who', 'whom', 'this', 'that', "that'll", 'these', 'tho  
se', 'am', 'is', 'are', 'was', 'were', 'be', 'been', 'being', 'have', 'has',  
'had', 'having', 'do', 'does', 'did', 'doing', 'a', 'an', 'the', 'and', 'bu  
t', 'if', 'or', 'because', 'as', 'until', 'while', 'of', 'at', 'by', 'for',  
'with', 'about', 'against', 'between', 'into', 'through', 'during', 'befor  
e', 'after', 'above', 'below', 'to', 'from', 'up', 'down', 'in', 'out', 'o  
n', 'off', 'over', 'under', 'again', 'further', 'then', 'once', 'here', 'the  
re', 'when', 'where', 'why', 'how', 'all', 'any', 'both', 'each', 'few', 'mo  
re', 'most', 'other', 'some', 'such', 'no', 'nor', 'not', 'only', 'own', 'sa  
me', 'so', 'than', 'too', 'very', 's', 't', 'can', 'will', 'just', 'don', "d  
on't", 'should', "should've", 'now', 'd', 'll', 'm', 'o', 're', 've', 'y',  
'ain', 'aren', "aren't", 'couldn', "couldn't", 'didn', "didn't", 'doesn', "d  
oesn't", 'hadn', "hadn't", 'hasn', "hasn't", 'haven', "haven't", 'isn', "is  
n't", 'ma', 'mightn', "mightn't", 'mustn', "mustn't", 'needn', "needn't", 's  
han', "shan't", 'shouldn', "shouldn't", 'wasn', "wasn't", 'weren', "were  
n't", 'won', "won't", 'wouldn', "wouldn't"]
```

In [5]:

```
1 print(word_tokens)
```

```
['Four', 'score', 'and', 'seven', 'years', 'ago', 'our', 'fathers', 'brough  
t', 'forth', 'on', 'this', 'continent', ',', ',', 'a', 'new', 'nation', ',', ',', 'con  
ceived', 'in', 'Liberty', ',', ',', 'and', 'dedicated', 'to', 'the', 'propositio  
n', 'that', 'all', 'men', 'are', 'created', 'equal', '.', 'Now', 'we', 'ar  
e', 'engaged', 'in', 'a', 'great', 'civil', 'war', ',', ',', 'testing', 'whethe  
r', 'that', 'nation', ',', ',', 'or', 'any', 'nation', 'so', 'conceived', 'and',  
'so', 'dedicated', ',', ',', 'can', 'long', 'endure', '.', 'We', 'are', 'met', 'o  
n', 'a', 'great', 'battle-field', 'of', 'that', 'war', '.', 'We', 'have', 'c  
ome', 'to', 'dedicate', 'a', 'portion', 'of', 'that', 'field', ',', ',', 'as',  
'a', 'final', 'resting', 'place', 'for', 'those', 'who', 'here', 'gave', 'th  
eir', 'lives', 'that', 'that', 'nation', 'might', 'live', '.', 'It', 'is',  
'altogether', 'fitting', 'and', 'proper', 'that', 'we', 'should', 'do', 'thi  
s', '.', 'But', ',', ',', 'in', 'a', 'larger', 'sense', ',', ',', 'we', 'can', 'not',  
'dedicate', '--', 'we', 'can', 'not', 'consecrate', '--', 'we', 'can', 'no  
t', 'hallow', '--', 'this', 'ground', '.', 'The', 'brave', 'men', ',', ',', 'livi  
ng', 'and', 'dead', ',', ',', 'who', 'struggled', 'here', ',', ',', 'have', 'consecrat  
ed', 'it', ',', ',', 'far', 'above', 'our', 'poor', 'power', 'to', 'add', 'or',  
'detract', '.', 'The', 'world', 'will', 'little', 'note', ',', ',', 'nor', 'lon  
g', 'remember', 'what', 'we', 'say', 'here', ',', ',', 'but', 'it', 'can', 'neve  
r', 'forget', 'what', 'they', 'did', 'here', '.', 'It', 'is', 'for', 'us',  
'the', 'living', ',', ',', 'rather', ',', ',', 'to', 'be', 'dedicated', 'here', 'to',  
'the', 'unfinished', 'work', 'which', 'they', 'who', 'fought', 'here', 'hav  
e', 'thus', 'far', 'so', 'nobly', 'advanced', '.', 'It', 'is', 'rather', 'fo  
r', 'us', 'to', 'be', 'here', 'dedicated', 'to', 'the', 'great', 'task', 're  
maining', 'before', 'us', '--', 'that', 'from', 'these', 'honored', 'dead',  
'we', 'take', 'increased', 'devotion', 'to', 'that', 'cause', 'for', 'whic  
h', 'they', 'gave', 'the', 'last', 'full', 'measure', 'of', 'devotion', '--  
, 'that', 'we', 'here', 'highly', 'resolve', 'that', 'these', 'dead', 'shal  
l', 'not', 'have', 'died', 'in', 'vain', '--', 'that', 'this', 'nation',  
, ',', 'under', 'God', ',', ',', 'shall', 'have', 'a', 'new', 'birth', 'of', 'freed  
om', '--', 'and', 'that', 'government', 'of', 'the', 'people', ',', ',', 'by', 't  
he', 'people', ',', ',', 'for', 'the', 'people', ',', ',', 'shall', 'not', 'perish',  
'from', 'the', 'earth', '.']
```

In [6]:

```
1 fil_sen = []  
2  
3 for w in word_tokens:  
4     if w not in stop_words:  
5         fil_sen.append(w)  
6
```

In [7]:

```
1 print(fil_sen)
```

```
['Four', 'score', 'seven', 'years', 'ago', 'fathers', 'brought', 'forth', 'c  
ontinent', ',', 'new', 'nation', ',', 'conceived', 'Liberty', ',', 'dedicate  
d', 'proposition', 'men', 'created', 'equal', '.', 'Now', 'engaged', 'grea  
t', 'civil', 'war', ',', 'testing', 'whether', 'nation', ',', 'nation', 'con  
ceived', 'dedicated', ',', 'long', 'endure', '.', 'We', 'met', 'great', 'bat  
tle-field', 'war', '.', 'We', 'come', 'dedicate', 'portion', 'field', ',',  
'final', 'resting', 'place', 'gave', 'lives', 'nation', 'might', 'live',  
'.', 'It', 'altogether', 'fitting', 'proper', '.', 'But', ',', 'larger', 'se  
nse', ',', 'dedicate', '--', 'consecrate', '--', 'hallow', '--', 'ground',  
'.', 'The', 'brave', 'men', ',', 'living', 'dead', ',', 'struggled', ',', 'c  
onsecrated', ',', 'far', 'poor', 'power', 'add', 'detract', '.', 'The', 'wor  
ld', 'little', 'note', ',', 'long', 'remember', 'say', ',', 'never', 'forge  
t', '.', 'It', 'us', 'living', ',', 'rather', ',', 'dedicated', 'unfinishe  
d', 'work', 'fought', 'thus', 'far', 'nobly', 'advanced', '.', 'It', 'rathe  
r', 'us', 'dedicated', 'great', 'task', 'remaining', 'us', '--', 'honored',  
'dead', 'take', 'increased', 'devotion', 'cause', 'gave', 'last', 'full', 'm  
easure', 'devotion', '--', 'highly', 'resolve', 'dead', 'shall', 'died', 'va  
in', '--', 'nation', ',', 'God', ',', 'shall', 'new', 'birth', 'freedom', '-  
-', 'government', 'people', ',', 'people', ',', 'people', ',', 'shall', 'per  
ish', 'earth', '.']
```

In [8]:

```
1 print(len(word_tokens))  
2 print(len(fil_sen))
```

```
310  
170
```

Extract and Display Bigrams

In [9]:

```
1 import nltk
2 bigrm = list(nltk.bigrams(word_tokens))
3 print(bigrm)
```

[('Four', 'score'), ('score', 'and'), ('and', 'seven'), ('seven', 'years'), ('years', 'ago'), ('ago', 'our'), ('our', 'fathers'), ('fathers', 'brought'), ('brought', 'forth'), ('forth', 'on'), ('on', 'this'), ('this', 'continent'), ('continent', ','), (',', 'a'), ('a', 'new'), ('new', 'nation'), ('nation', ','), (',', 'conceived'), ('conceived', 'in'), ('in', 'Liberty'), ('Liberty', ','), (',', 'and'), ('and', 'dedicated'), ('dedicated', 'to'), ('to', 'the'), ('the', 'proposition'), ('proposition', 'that'), ('that', 'all'), ('all', 'men'), ('men', 'are'), ('are', 'created'), ('created', 'equal'), ('equal', '.'), ('.', 'Now'), ('Now', 'we'), ('we', 'are'), ('are', 'engaged'), ('engaged', 'in'), ('in', 'a'), ('a', 'great'), ('great', 'civil'), ('civil', 'war'), ('war', ','), (',', 'testing'), ('testing', 'whether'), ('whether', 'that'), ('that', 'nation'), ('nation', ','), (',', 'or'), ('or', 'any'), ('any', 'nation'), ('nation', 'so'), ('so', 'conceived'), ('conceived', 'and'), ('and', 'so'), ('so', 'dedicated'), ('dedicated', ','), (',', 'can'), ('can', 'long'), ('long', 'endure'), ('endure', '.'), ('.', 'We'), ('We', 'are'), ('are', 'met'), ('met', 'on'), ('on', 'a'), ('a', 'great'), ('great', 'battle-field'), ('battle-field', 'of'), ('of', 'that'), ('that', 'war'), ('war', '.'), ('.', 'We'), ('We', 'have'), ('have', 'come'), ('come', 'to'), ('to', 'dedicate'), ('dedicate', 'a'), ('a', 'portion'), ('portion', 'of'), ('of', 'that'), ('that', 'field'), ('field', ','), (',', 'as'), ('as', 'a'), ('a', 'final'), ('final', 'resting'), ('resting', 'place'), ('place', 'for'), ('for', 'those'), ('those', 'who'), ('who', 'here'), ('here', 'gave'), ('gave', 'their'), ('their', 'lives'), ('lives', 'that'), ('that', 'that'), ('that', 'nation'), ('nation', 'might'), ('might', 'live'), ('live', '.'), ('.', 'It'), ('It', 'is'), ('is', 'altogether'), ('altogether', 'fitting'), ('fitting', 'and'), ('and', 'proper'), ('proper', 'that'), ('that', 'we'), ('we', 'should'), ('should', 'do'), ('do', 'this'), ('this', '.'), ('.', 'But'), ('But', ','), (',', 'in'), ('in', 'a'), ('a', 'larger'), ('larger', 'sense'), ('sense', ','), (',', 'we'), ('we', 'can'), ('can', 'not'), ('not', 'dedicate'), ('dedicate', '--'), ('--', 'we'), ('we', 'can'), ('can', 'not'), ('not', 'consecrate'), ('consecrate', '--'), ('--', 'we'), ('we', 'can'), ('can', 'not'), ('not', 'hallow'), ('hallow', '--'), ('--', 'this'), ('this', 'ground'), ('ground', '.'), ('.', 'The'), ('The', 'brave'), ('brave', 'men'), ('men', ','), (',', 'living'), ('living', 'and'), ('and', 'dead'), ('dead', ','), (',', 'who'), ('who', 'struggled'), ('struggled', 'here'), ('here', ','), (',', 'have'), ('have', 'consecrated'), ('consecrated', 'it'), ('it', ','), (',', 'far'), ('far', 'above'), ('above', 'our'), ('our', 'poor'), ('poor', 'power'), ('power', 'to'), ('to', 'add'), ('add', 'or'), ('or', 'detract'), ('detract', '.'), ('.', 'The'), ('The', 'world'), ('world', 'will'), ('will', 'little'), ('little', 'note'), ('note', ','), (',', 'nor'), ('nor', 'long'), ('long', 'remember'), ('remember', 'what'), ('what', 'we'), ('we', 'say'), ('say', 'here'), ('here', ','), (',', 'but'), ('but', 'it'), ('it', 'can'), ('can', 'never'), ('never', 'forget'), ('forget', 'what'), ('what', 'they'), ('they', 'did'), ('did', 'here'), ('here', '.'), ('.', 'It'), ('It', 'is'), ('is', 'for'), ('for', 'us'), ('us', 'the'), ('the', 'living'), ('living', ','), (',', 'rather'), ('rather', ','), (',', 'to'), ('to', 'be'), ('be', 'dedicated'), ('dedicated', 'here'), ('here', 'to'), ('to', 'the'), ('the', 'unfinished'), ('unfinished', 'work'), ('work', 'which'), ('which', 'they'), ('they', 'who'), ('who', 'fought'), ('fought', 'here'), ('here', 'have'), ('have', 'thus'), ('thus', 'far'), ('far', 'so'), ('so', 'nobly'), ('nobly', 'advanced'), ('advanced', '.'), ('.', 'It'), ('It', 'is'), ('is', 'rather'), ('rather', 'for'), ('for', 'us'), ('us', 'to'), ('to', 'be'), ('be', 'here'), ('here', 'dedicated'), ('dedicated', 'to'), ('to', 'the'), ('the', 'great'), ('great', 'task'), ('task', 'remaining'), ('remaining', 'before'), ('before', 'us'), ('u

```
s', '--'), ('--', 'that'), ('that', 'from'), ('from', 'these'), ('these', 'h
onored'), ('honored', 'dead'), ('dead', 'we'), ('we', 'take'), ('take', 'inc
reased'), ('increased', 'devotion'), ('devotion', 'to'), ('to', 'that'), ('t
hat', 'cause'), ('cause', 'for'), ('for', 'which'), ('which', 'they'), ('the
y', 'gave'), ('gave', 'the'), ('the', 'last'), ('last', 'full'), ('full', 'm
easure'), ('measure', 'of'), ('of', 'devotion'), ('devotion', '--'), ('--',
'that'), ('that', 'we'), ('we', 'here'), ('here', 'highly'), ('highly', 'res
olve'), ('resolve', 'that'), ('that', 'these'), ('these', 'dead'), ('dead',
'shall'), ('shall', 'not'), ('not', 'have'), ('have', 'died'), ('died', 'i
n'), ('in', 'vain'), ('vain', '--'), ('--', 'that'), ('that', 'this'), ('thi
s', 'nation'), ('nation', ','), (',', 'under'), ('under', 'God'), ('God',
','), (',', 'shall'), ('shall', 'have'), ('have', 'a'), ('a', 'new'), ('ne
w', 'birth'), ('birth', 'of'), ('of', 'freedom'), ('freedom', '--'), ('--',
'and'), ('and', 'that'), ('that', 'government'), ('government', 'of'), ('o
f', 'the'), ('the', 'people'), ('people', ','), (',', 'by'), ('by', 'the'),
('the', 'people'), ('people', ','), (',', 'for'), ('for', 'the'), ('the', 'p
eople'), ('people', ','), (',', 'shall'), ('shall', 'not'), ('not', 'peris
h'), ('perish', 'from'), ('from', 'the'), ('the', 'earth'), ('earth', '.')]

```

In [10]:

```
1 print(len(bigram))
```

309

From the Bigrams, Extract Collocations

In [11]:

```
1 from nltk.collocations import BigramCollocationFinder
2 from nltk.metrics import BigramAssocMeasures
3
4 finder = BigramCollocationFinder.from_words(bigram)
5 bigram_measures = BigramAssocMeasures.likelihood_ratio
6
7 print(finder.nbest(bigram_measures, 5))

```

```
[((('.', 'It'), ('It', 'is')), (('the', 'people'), ('people', ',')), (('we',
'can'), ('can', 'not'))), (('--', 'we'), ('we', 'can')), (('dedicated', 't
o'), ('to', 'the'))]
```

In [12]:

```
1 print(finder.nbest(bigram_measures, 50))
```

```
[(('.', 'It'), ('It', 'is')), (('the', 'people'), ('people', ', ')), (('we',  
'can'), ('can', 'not')), (('--', 'we'), ('we', 'can')), (('dedicated', 't  
o'), ('to', 'the')), ((' ', 'and'), ('and', 'dedicated')), ((' ', 'as'), ('a  
s', 'a')), ((' ', 'but'), ('but', 'it')), ((' ', 'by'), ('by', 'the')),  
((', ', 'can'), ('can', 'long')), ((' ', 'conceived'), ('conceived', 'in')),  
((', ', 'far'), ('far', 'above')), ((' ', 'for'), ('for', 'the')), ((' ', 'ha  
ve'), ('have', 'consecrated')), ((' ', 'living'), ('living', 'and')), ((' ',  
'nor'), ('nor', 'long')), ((' ', 'or'), ('or', 'any')), ((' ', 'rather'),  
('rather', ', ')), ((' ', 'testing'), ('testing', 'whether')), ((' ', 'unde  
r'), ('under', 'God')), ((' ', 'who'), ('who', 'struggled')), (('--', 'an  
d'), ('and', 'that')), (('--', 'this'), ('this', 'ground')), (('.', 'But'),  
('But', ', ')), (('.', 'Now'), ('Now', 'we')), (('But', ', '), (' ', 'in')),  
(( 'Four', 'score'), ('score', 'and')), (('Liberty', ', '), (' ', 'and')),  
(( 'Now', 'we'), ('we', 'are')), (('The', 'brave'), ('brave', 'men')), (('Th  
e', 'world'), ('world', 'will')), (('We', 'are'), ('are', 'met')), (('We',  
'have'), ('have', 'come')), (('a', 'final'), ('final', 'resting')), (('a',  
'larger'), ('larger', 'sense')), (('a', 'portion'), ('portion', 'of')), (('a  
bove', 'our'), ('our', 'poor')), (('add', 'or'), ('or', 'detract')), (('ag  
o', 'our'), ('our', 'fathers')), (('all', 'men'), ('men', 'are')), (('altoge  
ther', 'fitting'), ('fitting', 'and')), (('and', 'dead'), ('dead', ', ')),  
(( 'and', 'proper'), ('proper', 'that')), (('and', 'seven'), ('seven', 'year  
s')), (('and', 'so'), ('so', 'dedicated')), (('and', 'that'), ('that', 'gove  
rnment')), (('any', 'nation'), ('nation', 'so')), (('are', 'created'), ('cre  
ated', 'equal')), (('are', 'engaged'), ('engaged', 'in')), (('are', 'met'),  
('met', 'on'))]
```