Lab Assignment - 5

```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import regex
import nltk
from nltk.corpus import stopwords
```

```python
text= '''2001-02: Mike Denness incident, Kolkata Test, and breaking Bradman's record
During India's 2001 tour of South Africa, in the second Test match, referee Mike Denness fined four Indian players for excessive appealing, and fined

On the final day of the Kolkata Test against Australia in 2001, Tendulkar took three wickets, including the key wickets of Matthew Hayden and Adam Gi

In the 2002 series in the West Indies, Tendulkar started well, scoring 79 in the first Test. In the second Test at Port of Spain, Sachin Tendulkar sc

Then, in an unprecedented sequence, he scored just 0, 0, 8, and 0 in the next four innings.[130] He returned to form in the last Test scoring 41 and
```

```python
words = nltk.word_tokenize(text)
tags = nltk.pos_tag(words)
print(tags)
```

```
[('2001-02', 'CD'), (':', ':'), ('Mike', 'NNP'), ('Denness', 'NNP'), ('incident', 'NN'), (',', ','), ('Kolkata', 'NNP'), ('Test', 'NNP'), (',',
','), ('and', 'CC'), ('breaking', 'VBG'), ('Bradman', 'NNP'), ("'s", 'POS'), ('record', 'NN'), ('During', 'IN'), ('India', 'NNP'), ("'s", 'POS'),
('2001', 'CD'), ('tour', 'NN'), ('of', 'IN'), ('South', 'NNP'), ('Africa', 'NNP'), (',', ','), ('in', 'IN'), ('the', 'DT'), ('second', 'JJ'), ('Tes
t', 'NNP'), ('match', 'NN'), (',', ','), ('referee', 'VBP'), ('Mike', 'NNP'), ('Denness', 'NNP'), ('fined', 'VBD'), ('four', 'CD'), ('Indian', 'J
J'), ('players', 'NNS'), ('for', 'IN'), ('excessive', 'JJ'), ('appealing', 'NN'), (',', ','), ('and', 'CC'), ('fined', 'VBD'), ('the', 'DT'), ('Indi
an', 'JJ'), ('captain', 'NN'), ('Sourav', 'NNP'), ('Ganguly', 'NNP'), ('for', 'IN'), ('not', 'RB'), ('controlling', 'VBG'), ('his', 'PRP$'), ('tea
m', 'NN'), ('.', '.'), ('[', 'CC'), ('119', 'CD'), (']', 'JJ'), ('Denness', 'NNP'), ('suspended', 'VBD'), ('Tendulkar', 'NNP'), ('from', 'IN'), ('on
e', 'CD'), ('match', 'NN'), ('for', 'IN'), ('alleged', 'JJ'), ('ball', 'NN'), ('tampering', 'NN'), ('.', '.'), ('Television', 'NN'), ('cameras', 'NN
S'), ('captured', 'VBD'), ('images', 'NNS'), ('that', 'IN'), ('suggested', 'VBD'), ('Tendulkar', 'NNP'), ('may', 'MD'), ('have', 'VB'), ('been', 'VB
N'), ('involved', 'VBN'), ('in', 'IN'), ('cleaning', 'VBG'), ('the', 'DT'), ('seam', 'NN'), ('of', 'IN'), ('the', 'DT'), ('cricket', 'NN'), ('ball',
'NN'), ('.', '.'), ('[', 'CC'), ('120', 'CD'), (']', 'JJ'), ('[', '$'), ('121', 'CD'), (']', 'IN'), ('The', 'DT'), ('incident', 'NN'), ('escalated',
'VBD'), ('to', 'TO'), ('include', 'VB'), ('sports', 'NNS'), ('journalists', 'NNS'), ('accusing', 'VBG'), ('Denness', 'NNP'), ('of', 'IN'), ('racis
m', 'NN'), (',', ','), ('[', 'VBZ'), ('122', 'CD'), (']', 'NN'), ('and', 'CC'), ('led', 'VBD'), ('to', 'TO'), ('Denness', 'NNP'), ('being', 'VBG'),
('barred', 'VBN'), ('from', 'IN'), ('entering', 'VBG'), ('the', 'DT'), ('venue', 'NN'), ('of', 'IN'), ('the', 'DT'), ('third', 'JJ'), ('Test', 'NN
P'), ('match', 'NN'), ('.', '.'), ('The', 'DT'), ('ICC', 'NNP'), ('revoked', 'VBD'), ('the', 'DT'), ('status', 'NN'), ('of', 'IN'), ('the', 'DT'),
('match', 'NN'), ('as', 'IN'), ('a', 'DT'), ('Test', 'NN'), ('as', 'IN'), ('the', 'DT'), ('teams', 'NNS'), ('rejected', 'VBD'), ('the', 'DT'), ('app
ointed', 'VBN'), ('referee', 'NN'), ('.', '.'), ('[', 'CC'), ('123', 'CD'), (']', 'VBZ'), ('The', 'DT'), ('charges', 'NNS'), ('against', 'IN'), ('Te
ndulkar', 'NNP'), ('triggered', 'VBD'), ('a', 'DT'), ('massive', 'JJ'), ('backlash', 'NN'), ('from', 'IN'), ('the', 'DT'), ('Indian', 'JJ'), ('publi
c', 'NN'), ('.', '.'), ('[', 'CC'), ('124', 'CD'), (']', 'NN'), ('On', 'IN'), ('the', 'DT'), ('final', 'JJ'), ('day', 'NN'), ('of', 'IN'), ('the',
'DT'), ('Kolkata', 'NNP'), ('Test', 'NNP'), ('against', 'IN'), ('Australia', 'NNP'), ('in', 'IN'), ('2001', 'CD'), (',', ','), ('Tendulkar', 'NNP'),
('took', 'VBD'), ('three', 'CD'), ('wickets', 'NNS'), (',', ','), ('including', 'VBG'), ('the', 'DT'), ('key', 'JJ'), ('wickets', 'NNS'), ('of', 'I
N'), ('Matthew', 'NNP'), ('Hayden', 'NNP'), ('and', 'CC'), ('Adam', 'NNP'), ('Gilchrist', 'NNP'), (',', ','), ('who', 'WP'), ('were', 'VBD'), ('cent
urions', 'NNS'), ('in', 'IN'), ('the', 'DT'), ('previous', 'JJ'), ('Test', 'NNP'), ('.', '.'), ('His', 'PRP$'), ('three', 'CD'), ('wickets', 'NNS'),
('haul', 'VBP'), ('helped', 'VBD'), ('India', 'NNP'), ('win', 'VB'), ('the', 'DT'), ('match', 'NN'), ('.', '.'), ('[', 'CC'), ('125', 'CD'), (']',
'NN'), ('In', 'IN'), ('the', 'DT'), ('five-match', 'JJ'), ('ODI', 'NNP'), ('series', 'NN'), ('that', 'WDT'), ('followed', 'VBD'), (',', ','), ('he',
'PRP'), ('took', 'VBD'), ('his', 'PRP$'), ('100th', 'JJ'), ('wicket', 'NN'), ('in', 'IN'), ('ODIs', 'NNP'), (',', ','), ('claiming', 'VBG'), ('the',
'DT'), ('wicket', 'NN'), ('of', 'IN'), ('then', 'RB'), ('Australian', 'NNP'), ('captain', 'NN'), ('Steve', 'NNP'), ('Waugh', 'NNP'), ('in', 'IN'),
('the', 'DT'), ('final', 'JJ'), ('match', 'NN'), ('at', 'IN'), ('the', 'DT'), ('Fatorda', 'NNP'), ('Stadium', 'NNP'), ('in', 'IN'), ('Goa', 'NNP'),
('.', '.'), ('[', 'CC'), ('126', 'CD'), (']', 'NN'), ('In', 'IN'), ('the', 'DT'), ('2002', 'CD'), ('series', 'NN'), ('in', 'IN'), ('the', 'DT'), ('W
est', 'NNP'), ('Indies', 'NNPS'), (',', ','), ('Tendulkar', 'NNP'), ('started', 'VBD'), ('well', 'RB'), (',', ','), ('scoring', 'VBG'), ('79', 'C
D'), ('in', 'IN'), ('the', 'DT'), ('first', 'JJ'), ('Test', 'NNP'), ('.', '.'), ('In', 'IN'), ('the', 'DT'), ('second', 'JJ'), ('Test', 'NNP'), ('a
t', 'IN'), ('Port', 'NNP'), ('of', 'IN'), ('Spain', 'NNP'), (',', ','), ('Sachin', 'NNP'), ('Tendulkar', 'NNP'), ('scored', 'VBD'), ('117', 'CD'),
('in', 'IN'), ('the', 'DT'), ('first', 'JJ'), ('innings', 'NNS'), (',', ','), ('his', 'PRP$'), ('29th', 'CD'), ('Test', 'NNP'), ('century', 'NN'),
('in', 'IN'), ('his', 'PRP$'), ('93rd', 'CD'), ('Test', 'NNP'), ('match', 'NN'), (',', ','), ('to', 'TO'), ('equal', 'VB'), ('Donald', 'NNP'), ('Bra
dman', 'NNP'), ("'s", 'POS'), ('record', 'NN'), ('of', 'IN'), ('29', 'CD'), ('Test', 'NNP'), ('hundreds', 'NNS'), ('.', '.'), ('He', 'PRP'), ('was',
'VBD'), ('gifted', 'VBN'), ('a', 'DT'), ('Ferrari', 'NNP'), ('360', 'CD'), ('Modena', 'NNP'), ('by', 'IN'), ('Fiat', 'NNP'), ('through', 'IN'), ('Mi
chael', 'NNP'), ('Schumacher', 'NNP'), ('for', 'IN'), ('achieving', 'VBG'), ('this', 'DT'), ('feat', 'NN'), ('.', '.'), ('[', 'CC'), ('127', 'CD'),
(']', 'JJ'), ('[', '$'), ('128', 'CD'), (']', 'NNP'), ('[', 'VBD'), ('129', 'CD'), (']', 'NN'), ('Then', 'RB'), (',', ','), ('in', 'IN'), ('an', 'D
T'), ('unprecedented', 'JJ'), ('sequence', 'NN'), (',', ','), ('he', 'PRP'), ('scored', 'VBD'), ('just', 'RB'), ('0', 'CD'), (',', ','), ('0', 'C
D'), (',', ','), ('8', 'CD'), (',', ','), ('and', 'CC'), ('0', 'CD'), ('in', 'IN'), ('the', 'DT'), ('next', 'JJ'), ('four', 'CD'), ('innings', 'NN
S'), ('.', '.'), ('[', '$'), ('130', 'CD'), (']', 'NN'), ('He', 'PRP'), ('returned', 'VBD'), ('to', 'TO'), ('form', 'VB'), ('in', 'IN'), ('the', 'D
T'), ('last', 'JJ'), ('Test', 'NNP'), ('scoring', 'VBG'), ('41', 'CD'), ('and', 'CC'), ('86', 'CD'), (',', ','), ('one', 'CD'), ('half', 'JJ'), ('ce
ntury', 'NN'), ('.', '.'), ('However', 'RB'), (',', ','), ('India', 'NNP'), ('lost', 'VBD'), ('the', 'DT'), ('series', 'NN'), ('.', '.'), ('[', 'C
C'), ('131', 'CD'), (']', 'NN'), ('In', 'IN'), ('this', 'DT'), ('period', 'NN'), (',', ','), ('in', 'IN'), ('the', 'DT'), ('third', 'JJ'), ('Test',
'NNP'), ('match', 'NN'), ('against', 'IN'), ('England', 'NNP'), ('in', 'IN'), ('August', 'NNP'), ('2002', 'CD'), (',', ','), ('Tendulkar', 'NNP'),
('scored', 'VBD'), ('his', 'PRP$'), ('30th', 'JJ'), ('Test', 'NNP'), ('century', 'NN'), ('to', 'TO'), ('surpass', 'VB'), ('Bradman', 'NNP'), ("'s",
'POS'), ('haul', 'NN'), (',', ','), ('in', 'IN'), ('his', 'PRP$'), ('99th', 'CD'), ('Test', 'NNP'), ('match', 'NN'), ('.', '.'), ('[', 'CC'), ('13
2', 'CD'), (']', 'JJ'), ('[', '$'), ('133', 'CD'), (']', 'NN')]
```

```
In [17]: from nltk import bigrams
         bigrams = list(bigrams(words))
         print(bigrams)
```

```
[('2001-02', ':'), (':', 'Mike'), ('Mike', 'Denness'), ('Denness', 'incident'), ('incident', ','), (',', 'Kolkata'), ('Kolkata', 'Test'), ('Test',
','), (',', 'and'), ('and', 'breaking'), ('breaking', 'Bradman'), ('Bradman', "'s"), ("'s", 'record'), ('record', 'During'), ('During', 'India'),
('India', "'s"), ("'s", '2001'), ('2001', 'tour'), ('tour', 'of'), ('of', 'South'), ('South', 'Africa'), ('Africa', ','), (',', 'in'), ('in', 'th
e'), ('the', 'second'), ('second', 'Test'), ('Test', 'match'), ('match', ','), (',', 'referee'), ('referee', 'Mike'), ('Mike', 'Denness'), ('Dennes
s', 'fined'), ('fined', 'four'), ('four', 'Indian'), ('Indian', 'players'), ('players', 'for'), ('for', 'excessive'), ('excessive', 'appealing'),
('appealing', ','), (',', 'and'), ('and', 'fined'), ('fined', 'the'), ('the', 'Indian'), ('Indian', 'captain'), ('captain', 'Sourav'), ('Sourav', 'G
anguly'), ('Ganguly', 'for'), ('for', 'not'), ('not', 'controlling'), ('controlling', 'his'), ('his', 'team'), ('team', '.'), ('.', '['), ('[', '11
9'), ('119', ']'), (']', 'Denness'), ('Denness', 'suspended'), ('suspended', 'Tendulkar'), ('Tendulkar', 'from'), ('from', 'one'), ('one', 'match'),
('match', 'for'), ('for', 'alleged'), ('alleged', 'ball'), ('ball', 'tampering'), ('tampering', '.'), ('.', 'Television'), ('Television', 'camera
s'), ('cameras', 'captured'), ('captured', 'images'), ('images', 'that'), ('that', 'suggested'), ('suggested', 'Tendulkar'), ('Tendulkar', 'may'),
('may', 'have'), ('have', 'been'), ('been', 'involved'), ('involved', 'in'), ('in', 'cleaning'), ('cleaning', 'the'), ('the', 'seam'), ('seam', 'o
f'), ('of', 'the'), ('the', 'cricket'), ('cricket', 'ball'), ('ball', '.'), ('.', '['), ('[', '120'), ('120', ']'), (']', '['), ('[', '121'), ('12
1', ']'), (']', 'The'), ('The', 'incident'), ('incident', 'escalated'), ('escalated', 'to'), ('to', 'include'), ('include', 'sports'), ('sports', 'j
ournalists'), ('journalists', 'accusing'), ('accusing', 'Denness'), ('Denness', 'of'), ('of', 'racism'), ('racism', ','), (',', '['), ('[', '122'),
('122', ']'), (']', 'and'), ('and', 'led'), ('led', 'to'), ('to', 'Denness'), ('Denness', 'being'), ('being', 'barred'), ('barred', 'from'), ('fro
m', 'entering'), ('entering', 'the'), ('the', 'venue'), ('venue', 'of'), ('of', 'the'), ('the', 'third'), ('third', 'Test'), ('Test', 'match'), ('ma
tch', '.'), ('.', 'The'), ('The', 'ICC'), ('ICC', 'revoked'), ('revoked', 'the'), ('the', 'status'), ('status', 'of'), ('of', 'the'), ('the', 'matc
h'), ('match', 'as'), ('as', 'a'), ('a', 'Test'), ('Test', 'as'), ('as', 'the'), ('the', 'teams'), ('teams', 'rejected'), ('rejected', 'the'), ('th
e', 'appointed'), ('appointed', 'referee'), ('referee', '.'), ('.', '['), ('[', '123'), ('123', ']'), (']', 'The'), ('The', 'charges'), ('charges',
'against'), ('against', 'Tendulkar'), ('Tendulkar', 'triggered'), ('triggered', 'a'), ('a', 'massive'), ('massive', 'backlash'), ('backlash', 'fro
m'), ('from', 'the'), ('the', 'Indian'), ('Indian', 'public'), ('public', '.'), ('.', '['), ('[', '124'), ('124', ']'), (']', 'On'), ('On', 'the'),
('the', 'final'), ('final', 'day'), ('day', 'of'), ('of', 'the'), ('the', 'Kolkata'), ('Kolkata', 'Test'), ('Test', 'against'), ('against', 'Austral
ia'), ('Australia', 'in'), ('in', '2001'), ('2001', ','), (',', 'Tendulkar'), ('Tendulkar', 'took'), ('took', 'three'), ('three', 'wickets'), ('wick
ets', ','), (',', 'including'), ('including', 'the'), ('the', 'key'), ('key', 'wickets'), ('wickets', 'of'), ('of', 'Matthew'), ('Matthew', 'Hayde
n'), ('Hayden', 'and'), ('and', 'Adam'), ('Adam', 'Gilchrist'), ('Gilchrist', ','), (',', 'who'), ('who', 'were'), ('were', 'centurions'), ('centuri
ons', 'in'), ('in', 'the'), ('the', 'previous'), ('previous', 'Test'), ('Test', '.'), ('.', 'His'), ('His', 'three'), ('three', 'wickets'), ('wicket
s', 'haul'), ('haul', 'helped'), ('helped', 'India'), ('India', 'win'), ('win', 'the'), ('the', 'match'), ('match', '.'), ('.', '['), ('[', '125'),
('125', ']'), (']', 'In'), ('In', 'the'), ('the', 'five-match'), ('five-match', 'ODI'), ('ODI', 'series'), ('series', 'that'), ('that', 'followed'),
('followed', ','), (',', 'he'), ('he', 'took'), ('took', 'his'), ('his', '100th'), ('100th', 'wicket'), ('wicket', 'in'), ('in', 'ODIs'), ('ODIs',
','), (',', 'claiming'), ('claiming', 'the'), ('the', 'wicket'), ('wicket', 'of'), ('of', 'then'), ('then', 'Australian'), ('Australian', 'captai
n'), ('captain', 'Steve'), ('Steve', 'Waugh'), ('Waugh', 'in'), ('in', 'the'), ('the', 'final'), ('final', 'match'), ('match', 'at'), ('at', 'the'),
('the', 'Fatorda'), ('Fatorda', 'Stadium'), ('Stadium', 'in'), ('in', 'Goa'), ('Goa', '.'), ('.', '['), ('[', '126'), ('126', ']'), (']', 'In'), ('I
n', 'the'), ('the', '2002'), ('2002', 'series'), ('series', 'in'), ('in', 'the'), ('the', 'West'), ('West', 'Indies'), ('Indies', ','), (',', 'Tendu
lkar'), ('Tendulkar', 'started'), ('started', 'well'), ('well', ','), (',', 'scoring'), ('scoring', '79'), ('79', 'in'), ('in', 'the'), ('the', 'fir
st'), ('first', 'Test'), ('Test', '.'), ('.', 'In'), ('In', 'the'), ('the', 'second'), ('second', 'Test'), ('Test', 'at'), ('at', 'Port'), ('Port',
'of'), ('of', 'Spain'), ('Spain', ','), (',', 'Sachin'), ('Sachin', 'Tendulkar'), ('Tendulkar', 'scored'), ('scored', '117'), ('117', 'in'), ('in',
'the'), ('the', 'first'), ('first', 'innings'), ('innings', ','), (',', 'his'), ('his', '29th'), ('29th', 'Test'), ('Test', 'century'), ('century',
'in'), ('in', 'his'), ('his', '93rd'), ('93rd', 'Test'), ('Test', 'match'), ('match', ','), (',', 'to'), ('to', 'equal'), ('equal', 'Donald'), ('Don
ald', 'Bradman'), ('Bradman', "'s"), ("'s", 'record'), ('record', 'of'), ('of', '29'), ('29', 'Test'), ('Test', 'hundreds'), ('hundreds', '.'),
('.', 'He'), ('He', 'was'), ('was', 'gifted'), ('gifted', 'a'), ('a', 'Ferrari'), ('Ferrari', '360'), ('360', 'Modena'), ('Modena', 'by'), ('by', 'F
iat'), ('Fiat', 'through'), ('through', 'Michael'), ('Michael', 'Schumacher'), ('Schumacher', 'for'), ('for', 'achieving'), ('achieving', 'this'),
('this', 'feat'), ('feat', '.'), ('.', '['), ('[', '127'), ('127', ']'), (']', '['), ('[', '128'), ('128', ']'), (']', '['), ('[', '129'), ('129',
']'), (']', 'Then'), ('Then', ','), (',', 'in'), ('in', 'an'), ('an', 'unprecedented'), ('unprecedented', 'sequence'), ('sequence', ','), (',', 'h
e'), ('he', 'scored'), ('scored', 'just'), ('just', '0'), ('0', ','), (',', '0'), ('0', ','), (',', '8'), ('8', ','), (',', 'and'), ('and', '0'),
('0', 'in'), ('in', 'the'), ('the', 'next'), ('next', 'four'), ('four', 'innings'), ('innings', '.'), ('.', '['), ('[', '130'), ('130', ']'), (']',
'He'), ('He', 'returned'), ('returned', 'to'), ('to', 'form'), ('form', 'in'), ('in', 'the'), ('the', 'last'), ('last', 'Test'), ('Test', 'scorin
g'), ('scoring', '41'), ('41', 'and'), ('and', '86'), ('86', ','), (',', 'one'), ('one', 'half'), ('half', 'century'), ('century', '.'), ('.', 'Howe
ver'), ('However', ','), (',', 'India'), ('India', 'lost'), ('lost', 'the'), ('the', 'series'), ('series', '.'), ('.', '['), ('[', '131'),
(']', ']'), (']', 'In'), ('In', 'this'), ('this', 'period'), ('period', ','), (',', 'in'), ('in', 'the'), ('the', 'third'), ('third', 'Test'), ('Test', 'm
atch'), ('match', 'against'), ('against', 'England'), ('England', 'in'), ('in', 'August'), ('August', '2002'), ('2002', ','), (',', 'Tendulkar'),
('Tendulkar', 'scored'), ('scored', 'his'), ('his', '30th'), ('30th', 'Test'), ('Test', 'century'), ('century', 'to'), ('to', 'surpass'), ('surpas
s', 'Bradman'), ('Bradman', "'s"), ("'s", 'haul'), ('haul', ','), (',', 'in'), ('in', 'his'), ('his', '99th'), ('99th', 'Test'), ('Test', 'match'),
('match', '.'), ('.', '['), ('[', '132'), ('132', ']'), (']', '['), ('[', '133'), ('133', ']')]
```

```
In [18]: def filter_bigrams(tags):
             filtered_bigrams = []
             for i in range(len(tags) - 1):
                 first_word, first_tag = tags[i]
                 second_word, second_tag = tags[i + 1]

                 if (first_tag.startswith('JJ') and second_tag in ['NN', 'NNS']) \
                     or (first_tag in ['RB', 'RBR', 'RBS'] and second_tag.startswith('JJ') and not (second_tag == 'NN' or second_tag == 'NNS')) \
                     or (first_tag.startswith('JJ') and second_tag.startswith('JJ') and not (second_tag == 'NN' or second_tag == 'NNS')) \
                     or ((first_tag == 'NN' or first_tag == 'NNS') and second_tag.startswith('JJ') and not (second_tag == 'NN' or second_tag == 'NNS')) \
                     or (first_tag in ['RB', 'RBR', 'RBS'] and second_tag.startswith('VB') and second_tag != 'NNS'):
                     filtered_bigrams.append((first_word, second_word))
             return filtered_bigrams

         filtered_bigrams = filter_bigrams(tags)

         for bigram in filtered_bigrams:
           print(bigram)
```

```
('Indian', 'players')
('excessive', 'appealing')
('Indian', 'captain')
('not', 'controlling')
('alleged', 'ball')
('massive', 'backlash')
('Indian', 'public')
('final', 'day')
('key', 'wickets')
('100th', 'wicket')
('final', 'match')
('first', 'innings')
('unprecedented', 'sequence')
('half', 'century')
```

```
In [19]:  from nltk.probability import FreqDist
          from math import log

          def compute_pmi(filtered_bigrams, words):
            pmi_dict = {}
            word_freq = FreqDist(words)
            bigram_freq = FreqDist(filtered_bigrams)
            for bigram in filtered_bigrams:
              word1, word2 = bigram
              pmi = log((bigram_freq[bigram] * len(words)) / (word_freq[word1] * word_freq[word2]))
              pmi_dict[bigram] = pmi
            return pmi_dict

          pmi_dict = compute_pmi(filtered_bigrams, words)

          for bigram, pmi in pmi_dict.items():
            print(f"{bigram}: {pmi}")
```

```
('Indian', 'players'): 4.962844630259907
('excessive', 'appealing'): 6.061456918928017
('Indian', 'captain'): 4.269697449699962
('not', 'controlling'): 6.061456918928017
('alleged', 'ball'): 5.368309738368072
('massive', 'backlash'): 6.061456918928017
('Indian', 'public'): 4.962844630259907
('final', 'day'): 5.368309738368072
('key', 'wickets'): 4.962844630259907
('100th', 'wicket'): 5.368309738368072
('final', 'match'): 3.1710851610318525
('first', 'innings'): 4.675162557808126
('unprecedented', 'sequence'): 6.061456918928017
('half', 'century'): 4.962844630259907
```

```
In [20]:  positive_words = ["good", "great", "excellent", "amazing", "wonderful"]
          negative_words = ["not", "stopped", "difficult", "horrible", "Problem"]
          positive_bigrams = [bigram for bigram in filtered_bigrams if bigram[0] in positive_words or bigram[1] in positive_words]
          negative_bigrams = [bigram for bigram in filtered_bigrams if bigram[0] in negative_words or bigram[1] in negative_words]
          positive_pmi_dict = compute_pmi(positive_bigrams, words)
          negative_pmi_dict = compute_pmi(negative_bigrams, words)
          print("Positive Bigrams:")
          for bigram, pmi in positive_pmi_dict.items():
            print(f"{bigram}: {pmi}")

          print("\nNegative Bigrams:")
          for bigram, pmi in negative_pmi_dict.items():
            print(f"{bigram}: {pmi}")
```

```
Positive Bigrams:

Negative Bigrams:
('not', 'controlling'): 6.061456918928017
```

```
In [21]:  positive_avg_pmi = sum(pmi for pmi in positive_pmi_dict.values()) / 1
          negative_avg_pmi = sum(pmi for pmi in negative_pmi_dict.values()) / len(negative_pmi_dict)

          print("Average PMI for positive bigrams:", positive_avg_pmi)
          print("Average PMI for negative bigrams:", negative_avg_pmi)
          if positive_avg_pmi > negative_avg_pmi:
            print("The overall sentiment of the text is positive.")
          else:
            print("The overall sentiment of the text is negative.")
```

```
Average PMI for positive bigrams: 0.0
Average PMI for negative bigrams: 6.061456918928017
The overall sentiment of the text is negative.
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```