

In [ ]:

```
import pandas as pd
import nltk
nltk.download('punkt')
nltk.download('stopwords')
nltk.download('averaged_perceptron_tagger')
nltk.download('wordnet')
nltk.download('omw-1.4')
```

```
[nltk_data] Downloading package punkt to /Users/mehul/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package stopwords to /Users/mehul/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]   /Users/mehul/nltk_data...
[nltk_data]   Package averaged_perceptron_tagger is already up-to-
[nltk_data]   date!
[nltk_data] Downloading package wordnet to /Users/mehul/nltk_data...
[nltk_data]   Package wordnet is already up-to-date!
[nltk_data] Downloading package omw-1.4 to /Users/mehul/nltk_data...
[nltk_data]   Package omw-1.4 is already up-to-date!
```

Out[ ]:

True

In [ ]:

```
text_file = open("./sampletext.txt",'r')
data = text_file.read()
text_file.close() #good practice
```

### Tokenization

It is the process of breaking down a stream of text into words, phrases, symbols, or any other meaningful elements called tokens

In [ ]:

```
tokens = nltk.tokenize.word_tokenize(data)
print(tokens)
```

['Between', '2016', 'and', '2019', ',', 'the', 'state', 'forest', 'departmen  
t', 'under', 'the', 'BJP', 'government', 'had', 'launched', ',', 'Green', 'M  
aharashtra', ',', 'drive', 'with', 'an', 'aim', 'to', 'plant', '50', 'crore  
e', 'trees', 'across', 'the', 'state', 'in', 'the', 'four-year', 'period',  
, 'In', 'October', '2019', ',', 'the', 'government', 'had', 'claimed', 'i  
t', 'had', 'surpassed', 'the', 'target', 'by', 'planting', '33', 'crore', 't  
rees', 'in', 'July-September', '2019', '.', 'The', 'Indian', 'Express', 'ha  
d', 'found', 'that', 'non-forest', 'agencies', '-', 'such', 'as', 'gram', 'p  
anchayats', '-', 'which', 'were', 'tasked', 'with', 'planting', 'trees', 'ha  
d', 'not', 'uploaded', 'the', 'mandatory', 'audio-visual', 'proof', 'of', 't  
he', 'tree', 'plantation', 'drives', 'on', 'the', 'specially', 'created', 'p  
ortal', '.', 'In', 'Pune', 'Revenue', 'Division', ',', 'it', 'was', 'claime  
d', 'the', 'gram', 'panchayats', 'planted', '1.7', 'crore', 'saplings', ';',  
'however', ',', 'no', 'evidence', 'was', 'uploaded', 'for', '87', 'per', 'ce  
nt', '(', '1.49', 'crore', ')', 'saplings', '.', 'Also', ',', 'out', 'of',  
'the', '59', 'government', 'agencies', 'involved', 'in', 'the', 'drive', 'a  
s', 'many', 'as', '38', 'had', 'not', 'submitted', 'survival', 'reports', 'a  
bout', 'the', 'saplings', '.', 'This', 'year', ',', 'the', 'targets', 'set',  
'by', 'the', 'forest', 'department', 'were', 'comparatively', 'modest', '.',  
'For', 'example', ',', 'Pune', 'Circle', '-', 'which', 'comprises', 'three',  
'divisions', 'in', 'Pune', 'and', 'Solapur', 'district', '-', 'had', 'planne  
d', 'to', 'plant', '17', 'lakh', 'saplings', 'on', 'forest', 'land', '.', 'H  
owever', ',', 'it', 'may', 'not', 'meet', 'the', 'target', 'due', 'to', 'una  
vailability', 'of', 'funds', '.', 'Last', 'year', 'Pune', 'Circle', 'plante  
d', '70', 'lakh', 'saplings', 'on', 'forest', 'land', '.', 'In', 'Pune', 'Di  
vision', '-', 'which', 'comprises', 'six', 'talukas', 'namely', 'Maval',  
, 'Mulshi', 'Daund', 'Indapur', 'Baramati', 'and', 'Havvel  
i', '-', 'preparations', 'were', 'done', 'for', 'plantation', 'of', 'about',  
'4', 'lakh', 'trees', 'with', 'special', 'emphasis', 'on', 'teakwood', '.',  
'The', 'National', 'Forest', 'Policy', 'aims', 'and', 'emphasizes', 'at', 'm  
aintaining', '33', '%', 'of', 'the', 'country', 's', 'geographical', 'a  
rea', 'under', 'forest', 'and', 'green', 'cover', '.', 'In', 'view', 'of',  
'this', 'and', 'as', 'a', 'part', 'of', 'the', '50', 'crore', 'plantation',  
'programme', 'within', 'Maharashtra', ',', 'the', 'Maharashtra', 'Forest',  
'Department', 'aims', 'to', 'plant', '4', 'crore', 'saplings', 'all', 'ove  
r', 'the', 'State', 'between', '1st', 'to', 'July', '7th', ',', '2017', 't  
o', 'celebrate', ',', 'Vanmohotsav', ',', 'The', 'plantation', 'program  
me', ',', 'which', 'was', 'announced', 'in', '2016', 'with', 'the', 'aim',  
'of', 'planting', '2', 'crore', 'trees', 'on', '1st', 'July', '2016', 'was',  
'a', 'resounding', 'success', 'with', 'the', 'final', 'total', 'reported',  
'figure', 'of', '2.82', 'crore', 'saplings', 'planted', 'on', 'a', 'single',  
'day', '.', 'To', 'maintain', 'consistency', 'of', 'this', 'platform', 'with  
out', 'affecting', 'its', 'momentum', ',', 'the', 'Forest', 'Department', 'h  
as', 'set', 'the', 'target', 'of', 'plantation', 'of', '4', 'crore', ',', '1  
3', 'crore', 'and', '33', 'crore', 'saplings', 'under', 'the', 'mission', 'o  
f', '50', 'crore', 'plantation', 'which', 'shall', 'be', 'accomplished', 'i  
n', 'the', 'three', 'consecutive', 'years', 'viz', '.', '2017', ',', '2018',  
'and', '2019', '.', 'The', '4', 'crore', 'saplings', 'for', 'the', 'year',  
'2017', 'will', 'be', 'planted', 'during', 'the', 'Vanmohotsav', ',', 'Jul  
y', '1st', 'to', 'July', '7th', 'in', 'a', 'state-wide', 'drive', 'with', 't  
he', 'involvement', 'of', '33', 'government', 'departments', 'along', 'wit  
h', 'Students', 'of', 'Schools', 'and', 'Colleges', ',', 'NSS', ',', 'NCC',  
, ',', 'CSR', ',', 'NGOs', ',', 'Railways', ',', 'National', 'Highways', ',',  
'Defence', ',', 'NABARD', 'and', 'other', 'stakeholders', 'of', 'Society',  
, '.', 'In', 'a', 'first', 'of', 'its', 'kind', ',', 'a', '24-hour', 'toll',  
'free', 'helpline', 'number', '1926', 'called', ',', 'Hello', 'Forest', ',',  
'has', 'been', 'set', 'up', 'to', 'provide', 'information', 'regarding', 'pl  
antation', ',', 'protection', 'and', 'for', 'mass', 'awareness', '.', 'The',  
'Forest', 'Department', 'has', 'created', 'a', 'mobile', 'application', 'cal  
led', ',', 'My', 'Plants', ',', 'to', 'record', 'details', 'of', 'the', 'pla  
ntation', 'such', 'as', 'numbers', ',', 'species', 'and', 'location', 'int  
o', 'the', 'computer', 'system', 'of', 'the', 'Forest', 'Department', '.',  
'All', 'volunteers', 'at', 'individual', ',', 'collective', 'and', 'organiza  
tional', 'level', 'should', 'download', 'and', 'use', 'this', 'application',

'to', 'record', 'their', 'tree', 'plantation', 'work', 'through', 'the', 'ap  
plication', ',', 'which', 'will', 'be', 'operational', 'from', '1st', 'Jul  
y', 'to', '7th', 'July', '.', 'In', 'consonance', 'of', 'the', 'public', 'pa  
rticipation', ',', 'the', 'Maharashtra', 'Forest', 'Department', 'has', 'ini  
tiated', 'the', ',', 'Maharashtra', 'Harit', 'Sena', ',', '/', ',', 'Green',  
'Army', ',', 'which', 'is', 'a', 'body', 'of', 'dedicated', 'volunteers', 't  
o', 'participate', 'in', 'the', 'plantation', ',', 'protection', ',', 'and',  
'activities', 'in', 'forest', ',', 'wildlife', ',', 'and', 'related', 'secto  
rs', 'around', 'the', 'year', '.', 'Individuals', 'and', 'organisations', 'i  
nterested', 'in', 'volunteering', 'can', 'register', 'on', 'the', 'Green',  
'Army', 'website', 'www.greenarmy.mahaforest.gov.in', 'An', 'integrated', 'd  
rive', 'has', 'been', 'set', 'in', 'place', 'to', 'ensure', 'seamless', 'an  
d', 'successful', 'participation', 'from', 'all', 'stakeholders', 'of', 'soc  
iety', ',', 'especially', 'the', 'public', '']

### Stop words

Text and document classification over social media, such as Twitter, Facebook, and so on is usually affected by the noisy nature (abbreviations, irregular forms) of the text corpuses. eg: and, the, a etc

In [ ]:

```
from nltk.corpus import stopwords
stop_words = set(stopwords.words('english'))
# filtered_sentence = [w for w in tokens if not w in stop_words]
filtered_sentence = []

for w in tokens:
    if w not in stop_words:
        filtered_sentence.append(w)

print("Tokens Below \n")
print(tokens)
print("\nFiltered Sentence Below \n")
print(filtered_sentence)
```

## Tokens Below

['Between', '2016', 'and', '2019', ',', 'the', 'state', 'forest', 'departmen  
t', 'under', 'the', 'BJP', 'government', 'had', 'launched', ',', 'Green', 'M  
aharashtra', ',', 'drive', 'with', 'an', 'aim', 'to', 'plant', '50', 'crore  
e', 'trees', 'across', 'the', 'state', 'in', 'the', 'four-year', 'period',  
, 'In', 'October', '2019', ',', 'the', 'government', 'had', 'claimed', 'i  
t', 'had', 'surpassed', 'the', 'target', 'by', 'planting', '33', 'crore', 't  
rees', 'in', 'July-September', '2019', '.', 'The', 'Indian', 'Express', 'ha  
d', 'found', 'that', 'non-forest', 'agencies', '-', 'such', 'as', 'gram', 'p  
anchayats', '-', 'which', 'were', 'tasked', 'with', 'planting', 'trees', 'ha  
d', 'not', 'uploaded', 'the', 'mandatory', 'audio-visual', 'proof', 'of', 't  
he', 'tree', 'plantation', 'drives', 'on', 'the', 'specially', 'created', 'p  
ortal', '.', 'In', 'Pune', 'Revenue', 'Division', ',', 'it', 'was', 'claime  
d', 'the', 'gram', 'panchayats', 'planted', '1.7', 'crore', 'saplings', ';',  
'however', ',', 'no', 'evidence', 'was', 'uploaded', 'for', '87', 'per', 'ce  
nt', '(', '1.49', 'crore', ')', 'saplings', '.', 'Also', ',', 'out', 'of',  
'the', '59', 'government', 'agencies', 'involved', 'in', 'the', 'drive', 'a  
s', 'many', 'as', '38', 'had', 'not', 'submitted', 'survival', 'reports', 'a  
bout', 'the', 'saplings', '.', 'This', 'year', ',', 'the', 'targets', 'set',  
'by', 'the', 'forest', 'department', 'were', 'comparatively', 'modest', '.',  
'For', 'example', ',', 'Pune', 'Circle', '-', 'which', 'comprises', 'three',  
'divisions', 'in', 'Pune', 'and', 'Solapur', 'district', '-', 'had', 'planne  
d', 'to', 'plant', '17', 'lakh', 'saplings', 'on', 'forest', 'land', '.', 'H  
owever', ',', 'it', 'may', 'not', 'meet', 'the', 'target', 'due', 'to', 'una  
vailability', 'of', 'funds', '.', 'Last', 'year', 'Pune', 'Circle', 'plante  
d', '70', 'lakh', 'saplings', 'on', 'forest', 'land', '.', 'In', 'Pune', 'Di  
vision', '-', 'which', 'comprises', 'six', 'talukas', 'namely', 'Maval',  
, 'Mulshi', 'Daund', 'Indapur', 'Baramati', 'and', 'Havvel  
i', '-', 'preparations', 'were', 'done', 'for', 'plantation', 'of', 'about',  
'4', 'lakh', 'trees', 'with', 'special', 'emphasis', 'on', 'teakwood', '.',  
'The', 'National', 'Forest', 'Policy', 'aims', 'and', 'emphasizes', 'at', 'm  
aintaining', '33', '%', 'of', 'the', 'country', 's', 'geographical', 'a  
rea', 'under', 'forest', 'and', 'green', 'cover', '.', 'In', 'view', 'of',  
'this', 'and', 'as', 'a', 'part', 'of', 'the', '50', 'crore', 'plantation',  
'programme', 'within', 'Maharashtra', ',', 'the', 'Maharashtra', 'Forest',  
'Department', 'aims', 'to', 'plant', '4', 'crore', 'saplings', 'all', 'ove  
r', 'the', 'State', 'between', '1st', 'to', 'July', '7th', ',', '2017', 't  
o', 'celebrate', ',', 'Vanmohotsav', ',', 'The', 'plantation', 'program  
me', ',', 'which', 'was', 'announced', 'in', '2016', 'with', 'the', 'aim',  
'of', 'planting', '2', 'crore', 'trees', 'on', '1st', 'July', '2016', 'was',  
'a', 'resounding', 'success', 'with', 'the', 'final', 'total', 'reported',  
'figure', 'of', '2.82', 'crore', 'saplings', 'planted', 'on', 'a', 'single',  
'day', '.', 'To', 'maintain', 'consistency', 'of', 'this', 'platform', 'with  
out', 'affecting', 'its', 'momentum', ',', 'the', 'Forest', 'Department', 'h  
as', 'set', 'the', 'target', 'of', 'plantation', 'of', '4', 'crore', ',', '1  
3', 'crore', 'and', '33', 'crore', 'saplings', 'under', 'the', 'mission', 'o  
f', '50', 'crore', 'plantation', 'which', 'shall', 'be', 'accomplished', 'i  
n', 'the', 'three', 'consecutive', 'years', 'viz', '.', '2017', ',', '2018',  
'and', '2019', '.', 'The', '4', 'crore', 'saplings', 'for', 'the', 'year',  
'2017', 'will', 'be', 'planted', 'during', 'the', 'Vanmohotsav', ',', 'Jul  
y', '1st', 'to', 'July', '7th', 'in', 'a', 'state-wide', 'drive', 'with', 't  
he', 'involvement', 'of', '33', 'government', 'departments', 'along', 'wit  
h', 'Students', 'of', 'Schools', 'and', 'Colleges', ',', 'NSS', ',', 'NCC',  
, 'CSR', ',', 'NGOs', ',', 'Railways', ',', 'National', 'Highways', ',',  
'Defence', ',', 'NABARD', 'and', 'other', 'stakeholders', 'of', 'Society',  
, 'In', 'a', 'first', 'of', 'its', 'kind', ',', 'a', '24-hour', 'toll',  
'free', 'helpline', 'number', '1926', 'called', ',', 'Hello', 'Forest', ',',  
'has', 'been', 'set', 'up', 'to', 'provide', 'information', 'regarding', 'pl  
antation', ',', 'protection', 'and', 'for', 'mass', 'awareness', '.', 'The',  
'Forest', 'Department', 'has', 'created', 'a', 'mobile', 'application', 'cal  
led', ',', 'My', 'Plants', ',', 'to', 'record', 'details', 'of', 'the', 'pla  
ntation', 'such', 'as', 'numbers', ',', 'species', 'and', 'location', 'int  
o', 'the', 'computer', 'system', 'of', 'the', 'Forest', 'Department', '.',

'All', 'volunteers', 'at', 'individual', ',', 'collective', 'and', 'organizational', 'level', 'should', 'download', 'and', 'use', 'this', 'application', 'to', 'record', 'their', 'tree', 'plantation', 'work', 'through', 'the', 'application', ',', 'which', 'will', 'be', 'operational', 'from', '1st', 'July', 'to', '7th', 'July', '., 'In', 'consonance', 'of', 'the', 'public', 'participation', ',', 'the', 'Maharashtra', 'Forest', 'Department', 'has', 'initiated', 'the', ',', 'Maharashtra', 'Harit', 'Sena', ',', '/', ',', 'Green', 'Army', ',', 'which', 'is', 'a', 'body', 'of', 'dedicated', 'volunteers', 'to', 'participate', 'in', 'the', 'plantation', ',', 'protection', ',', 'and', 'activities', 'in', 'forest', ',', 'wildlife', ',', 'and', 'related', 'sectors', 'around', 'the', 'year', '., 'Individuals', 'and', 'organisations', 'interested', 'in', 'volunteering', 'can', 'register', 'on', 'the', 'Green', 'Army', 'website', 'www.greenarmy.mahaforest.gov.in', 'An', 'integrated', 'drive', 'has', 'been', 'set', 'in', 'place', 'to', 'ensure', 'seamless', 'and', 'successful', 'participation', 'from', 'all', 'stakeholders', 'of', 'society', ',', 'especially', 'the', 'public', '., ']

#### Filtered Sentence Below

[ 'Between', '2016', '2019', ',', 'state', 'forest', 'department', 'BJP', 'government', 'launched', ',', 'Green', 'Maharashtra', ',', 'drive', 'aim', 'plant', '50', 'crore', 'trees', 'across', 'state', 'four-year', 'period', '., 'In', 'October', '2019', ',', 'government', 'claimed', 'surpassed', 'target', 'planting', '33', 'crore', 'trees', 'July-September', '2019', '., 'The', 'Indian', 'Express', 'found', 'non-forest', 'agencies', '-', 'gram', 'panchayats', '-', 'tasked', 'planting', 'trees', 'uploaded', 'mandatory', 'audio-visual', 'proof', 'tree', 'plantation', 'drives', 'specially', 'created', 'portal', '., 'In', 'Pune', 'Revenue', 'Division', ',', 'claimed', 'gram', 'panchayats', 'planted', '1.7', 'crore', 'saplings', ';', 'however', ',', 'evidence', 'uploaded', '87', 'per', 'cent', '(', '1.49', 'crore', ')', 'saplings', '., 'Also', ',', '59', 'government', 'agencies', 'involved', 'drive', 'many', '38', 'submitted', 'survival', 'reports', 'saplings', '., 'This', 'year', ',', 'targets', 'set', 'forest', 'department', 'comparatively', 'modest', '., 'For', 'example', ',', 'Pune', 'Circle', '-', 'comprises', 'three', 'divisions', 'Pune', 'Solapur', 'district', '-', 'planned', 'plant', '17', 'lakh', 'saplings', 'forest', 'land', '., 'However', ',', 'may', 'meet', 'target', 'due', 'unavailability', 'funds', '., 'Last', 'year', 'Pune', 'Circle', 'planted', '70', 'lakh', 'saplings', 'forest', 'land', '., 'In', 'Pune', 'Division', '-', 'comprises', 'six', 'talukas', 'namely', 'Maval', ',', 'Mulshi', ',', 'Daund', ',', 'Indapur', ',', 'Baramati', 'Havveli', '-', 'preparations', 'done', 'plantation', '4', 'lakh', 'trees', 'special', 'emphasis', 'teakwood', '., 'The', 'National', 'Forest', 'Policy', 'aims', 'emphasizes', 'maintaining', '33', '%', 'country', ',', 'geographical', 'area', 'forest', 'green', 'cover', '., 'In', 'view', 'part', '50', 'crore', 'plantation', 'programme', 'within', 'Maharashtra', ',', 'Maharashtra', 'Forest', 'Department', 'aims', 'plant', '4', 'crore', 'saplings', 'State', '1st', 'July', '7th', ',', '2017', 'celebrate', ',', 'Vanmohotsav', ',', '., 'The', 'plantation', 'programme', ',', 'announced', '2016', 'aim', 'planting', '2', 'crore', 'trees', '1st', 'July', '2016', 'resounding', 'success', 'final', 'total', 'reported', 'figure', '2.82', 'crore', 'saplings', 'planted', 'single', 'day', '., 'To', 'maintain', 'consistency', 'platform', 'without', 'affecting', 'momentum', ',', 'Forest', 'Department', 'set', 'target', 'plantation', '4', 'crore', ',', '13', 'crore', '33', 'crore', 'saplings', 'mission', '50', 'crore', 'plantation', 'shall', 'accomplished', 'three', 'consecutive', 'years', 'viz', '., '2017', ',', '2018', '2019', '., 'The', '4', 'crore', 'saplings', 'year', '2017', 'planted', 'Vanmohotsav', ',', 'July', '1st', 'July', '7th', 'state-wide', 'drive', 'involvement', '33', 'government', 'departments', 'along', 'Students', 'Schools', 'Colleges', ',', 'NS S', ',', 'NCC', ',', 'CSR', ',', 'NGOs', ',', 'Railways', ',', 'National', 'Highways', ',', 'Defence', ',', 'NABARD', 'stakeholders', 'Society', '., 'In', 'first', 'kind', ',', '24-hour', 'toll', 'free', 'helpline', 'number', '1926', 'called', ',', 'Hello', 'Forest', ',', 'set', 'provide', 'information', 'regarding', 'plantation', ',', 'protection', 'mass', 'awareness', '., 'The', 'Forest', 'Department', 'created', 'mobile', 'application', 'called',

```
''', 'My', 'Plants', '', 'record', 'details', 'plantation', 'numbers', '',  
'species', 'location', 'computer', 'system', 'Forest', 'Department', '.', 'A  
ll', 'volunteers', 'individual', '', 'collective', 'organizational', 'leve  
l', 'download', 'use', 'application', 'record', 'tree', 'plantation', 'wor  
k', 'application', '', 'operational', '1st', 'July', '7th', 'July', '.', 'I  
n', 'consonance', 'public', 'participation', '', 'Maharashtra', 'Forest',  
'Department', 'initiated', '', 'Maharashtra', 'Harit', 'Sena', '', '/',  
'', 'Green', 'Army', '', 'body', 'dedicated', 'volunteers', 'participate',  
'plantation', '', 'protection', '', 'activities', 'forest', '', 'wildlif  
e', '', 'related', 'sectors', 'around', 'year', '.', 'Individuals', 'organi  
sations', 'interested', 'volunteering', 'register', 'Green', 'Army', 'websit  
e', 'www.greenarmy.mahaforest.gov.in', 'An', 'integrated', 'drive', 'set',  
'place', 'ensure', 'seamless', 'successful', 'participation', 'stakeholder  
s', 'society', '', 'especially', 'public', '.']
```

## POS Tagging

Part-of-speech (POS) tagging is a popular Natural Language Processing process which refers to categorizing words in a text (corpus) in correspondence with a particular part of speech, depending on the definition of the word and its context

```
In [ ]: from nltk import pos_tag  
pos = pos_tag(filtered_sentence)  
print(pos)
```

[('Between', 'IN'), ('2016', 'CD'), ('2019', 'CD'), (',', ','), ('state', 'NN'), ('forest', 'JJS'), ('department', 'NN'), ('BJP', 'NNP'), ('government', 'NN'), ('launched', 'VBD'), (',', 'NNP'), ('Green', 'NNP'), ('Maharashtra', 'NNP'), (',', 'NNP'), ('drive', 'NN'), ('aim', 'NN'), ('plant', 'NN'), ('50', 'CD'), ('crore', 'NN'), ('trees', 'NNS'), ('across', 'IN'), ('state', 'NN'), ('four-year', 'JJ'), ('period', 'NN'), ('.', '.'), ('In', 'IN'), ('October', 'NNP'), ('2019', 'CD'), (',', ','), ('government', 'NN'), ('claimed', 'VBD'), ('surpassed', 'JJ'), ('target', 'NN'), ('planting', 'VBG'), ('33', 'CD'), ('crore', 'NN'), ('trees', 'NNS'), ('July-September', 'NNP'), ('2019', 'CD'), ('.', '.'), ('The', 'DT'), ('Indian', 'JJ'), ('Express', 'NNP'), ('found', 'VBD'), ('non-forest', 'JJS'), ('agencies', 'NNS'), ('-', 'VBP'), ('gram', 'JJ'), ('panchayats', 'NNS'), ('-', 'VBP'), ('tasked', 'VBN'), ('planting', 'NN'), ('trees', 'NNS'), ('uploaded', 'JJ'), ('mandatory', 'JJ'), ('audio-visual', 'JJ'), ('proof', 'NN'), ('tree', 'JJ'), ('plantation', 'NN'), ('drives', 'NNS'), ('specially', 'RB'), ('created', 'VBD'), ('portal', 'JJ'), ('.', '.'), ('In', 'IN'), ('Pune', 'NNP'), ('Revenue', 'NNP'), ('Division', 'NNP'), (',', ','), ('claimed', 'VBD'), ('gram', 'JJ'), ('panchayats', 'NNS'), ('planted', 'VBD'), ('1.7', 'CD'), ('crore', 'NN'), ('saplings', 'NNS'), (',', ':'), ('however', 'RB'), (',', ','), ('evidence', 'NN'), ('uploaded', 'VBD'), ('87', 'CD'), ('per', 'IN'), ('cent', 'NN'), ('.', '.'), ('1.49', 'CD'), ('crore', 'NN'), (',', ','), ('saplings', 'NNS'), ('.', '.'), ('Also', 'RB'), (',', ','), ('59', 'CD'), ('government', 'NN'), ('agencies', 'NNS'), ('involved', 'VBD'), ('drive', 'JJ'), ('many', 'JJ'), ('38', 'CD'), ('submitted', 'VBD'), ('survival', 'NN'), ('reports', 'NNS'), ('saplings', 'NNS'), ('.', '.'), ('This', 'DT'), ('year', 'NN'), (',', ','), ('targets', 'NNS'), ('set', 'VBP'), ('forest', 'JJS'), ('department', 'NN'), ('comparatively', 'RB'), ('modest', 'JJ'), ('.', '.'), ('For', 'IN'), ('example', 'NN'), (',', ','), ('Pune', 'NNP'), ('Circle', 'NNP'), ('-', 'NNP'), ('comprises', 'VBZ'), ('three', 'CD'), ('divisions', 'NNS'), ('Pune', 'NNP'), ('Solapur', 'NNP'), ('district', 'NN'), ('-', 'NN'), ('planned', 'VBD'), ('plant', 'NN'), ('17', 'CD'), ('lakh', 'JJ'), ('saplings', 'NNS'), ('forest', 'JJS'), ('land', 'NN'), ('.', '.'), ('However', 'RB'), (',', ','), ('may', 'MD'), ('meet', 'VB'), ('target', 'NN'), ('due', 'JJ'), ('unavailability', 'NN'), ('funds', 'NNS'), ('.', '.'), ('Last', 'JJ'), ('year', 'NN'), ('Pune', 'NNP'), ('Circle', 'NNP'), ('planted', 'VBD'), ('70', 'CD'), ('lakh', 'JJ'), ('saplings', 'NNS'), ('forest', 'JJS'), ('land', 'NN'), ('.', '.'), ('In', 'IN'), ('Pune', 'NNP'), ('Division', 'NNP'), ('-', 'NNP'), ('comprises', 'VBZ'), ('six', 'CD'), ('talukas', 'NN'), ('namely', 'RB'), ('Maval', 'NNP'), (',', ','), ('Mulshi', 'NNP'), (',', ','), ('Daund', 'NNP'), (',', ','), ('Indapur', 'NNP'), (',', ','), ('Baramati', 'NNP'), ('Havveli', 'NNP'), ('-', 'NNP'), ('preparations', 'NNS'), ('done', 'VBN'), ('plantation', 'NN'), ('4', 'CD'), ('lakh', 'NN'), ('trees', 'NNS'), ('special', 'JJ'), ('emphasis', 'NN'), ('teakwood', 'NN'), ('.', '.'), ('The', 'DT'), ('National', 'NNP'), ('Forest', 'NNP'), ('Policy', 'NNP'), ('aims', 'VBZ'), ('emphasizes', 'VBZ'), ('maintaining', 'VBG'), ('33', 'CD'), ('%', 'NN'), ('country', 'NN'), (',', 'NNP'), ('geographical', 'JJ'), ('area', 'NN'), ('forest', 'JJS'), ('green', 'JJ'), ('cover', 'NN'), ('.', '.'), ('In', 'IN'), ('view', 'NN'), ('part', 'NN'), ('50', 'CD'), ('crore', 'NN'), ('plantation', 'NN'), ('programme', 'NN'), ('within', 'IN'), ('Maharashtra', 'NNP'), (',', ','), ('Maharashtra', 'NNP'), ('Forest', 'NNP'), ('Department', 'NNP'), ('aims', 'VBZ'), ('plant', 'NN'), ('4', 'CD'), ('crore', 'NN'), ('saplings', 'NNS'), ('State', 'NNP'), ('1st', 'CD'), ('July', 'NNP'), ('7th', 'CD'), (',', ','), ('2017', 'CD'), ('celebrate', 'NN'), (',', 'NNP'), ('Vanmohotsav', 'NNP'), (',', 'NNP'), ('.', '.'), ('The', 'DT'), ('plantation', 'NN'), ('programme', 'NN'), (',', ','), ('announced', 'VBD'), ('2016', 'CD'), ('aim', 'NN'), ('planting', 'VBG'), ('2', 'CD'), ('crore', 'NN'), ('trees', 'NNS'), ('1st', 'CD'), ('July', 'NNP'), ('2016', 'CD'), ('resounding', 'VBG'), ('success', 'NN'), ('final', 'JJ'), ('total', 'NN'), ('reported', 'VBD'), ('figure', 'NN'), ('2.82', 'CD'), ('crore', 'NN'), ('saplings', 'NNS'), ('planted', 'VBD'), ('single', 'JJ'), ('day', 'NN'), ('.', '.'), ('To', 'TO'), ('maintain', 'VB'), ('consistency', 'NN'), ('platform', 'NN'), ('without', 'IN'), ('affecting', 'VBG'), ('momentum', 'NN'), (',', ','), ('Forest', 'NNP'), ('Department', 'NNP'), ('set', 'VBD'), ('target', 'NN'), ('plantation', 'NN'), ('4', 'CD'), ('crore', 'NN'), (',', ','), ('13', 'CD'), ('crore', 'NN'), ('33', 'C

D'), ('crore', 'NN'), ('saplings', 'NNS'), ('mission', 'NN'), ('50', 'CD'), ('crore', 'NN'), ('plantation', 'NN'), ('shall', 'MD'), ('accomplished', 'VB'), ('three', 'CD'), ('consecutive', 'JJ'), ('years', 'NNS'), ('viz', 'R'), ('.', '.'), ('2017', 'CD'), ('.', '.'), ('2018', 'CD'), ('2019', 'CD'), ('.', '.'), ('The', 'DT'), ('4', 'CD'), ('crore', 'NN'), ('saplings', 'NN'), ('year', 'NN'), ('2017', 'CD'), ('planted', 'VBD'), ('Vanmohotsav', 'NN'), ('July', 'NNP'), ('1st', 'CD'), ('July', 'NNP'), ('7th', 'CD'), ('state-wide', 'JJ'), ('drive', 'NN'), ('involvement', 'NN'), ('33', 'CD'), ('government', 'NN'), ('departments', 'NNS'), ('along', 'IN'), ('Students', 'NNS'), ('Schools', 'NNP'), ('Colleges', 'NNP'), ('.', '.'), ('NSS', 'NNP'), ('NCC', 'NNP'), ('.', '.'), ('CSR', 'NNP'), ('.', '.'), ('NGOs', 'NNP'), ('.', '.'), ('Railways', 'NNP'), ('.', '.'), ('National', 'NNP'), ('Highways', 'NNP'), ('.', '.'), ('Defence', 'NNP'), ('.', '.'), ('NABARD', 'NNP'), ('stakeholders', 'VBZ'), ('Society', 'NNP'), ('.', '.'), ('In', 'IN'), ('first', 'JJ'), ('kind', 'NN'), ('.', '.'), ('24-hour', 'JJ'), ('toll', 'NN'), ('free', 'JJ'), ('helpline', 'NN'), ('number', 'NN'), ('1926', 'CD'), ('called', 'VBD'), ('.', '.'), ('Hello', 'NNP'), ('Forest', 'NNP'), ('set', 'VBD'), ('provide', 'JJ'), ('information', 'NN'), ('regarding', 'VBG'), ('plantation', 'NN'), ('.', '.'), ('protection', 'NN'), ('mass', 'NN'), ('awareness', 'NN'), ('.', '.'), ('The', 'DT'), ('Forest', 'NNP'), ('Department', 'NNP'), ('created', 'VBD'), ('mobile', 'JJ'), ('application', 'NN'), ('called', 'VBN'), ('.', '.'), ('My', 'PRP\$'), ('Plants', 'NNS'), ('.', '.'), ('record', 'NN'), ('details', 'NNS'), ('plantation', 'NN'), ('numbers', 'NNS'), ('.', '.'), ('species', 'NNS'), ('location', 'VBP'), ('computer', 'NN'), ('system', 'NN'), ('Forest', 'NNP'), ('Department', 'NNP'), ('.', '.'), ('All', 'NNP'), ('volunteers', 'NNS'), ('individual', 'JJ'), ('.', '.'), ('collective', 'JJ'), ('organizational', 'JJ'), ('level', 'NN'), ('download', 'NN'), ('use', 'NN'), ('application', 'NN'), ('record', 'NN'), ('tree', 'NN'), ('plantation', 'NN'), ('work', 'NN'), ('application', 'NN'), ('.', '.'), ('operational', 'JJ'), ('1st', 'CD'), ('July', 'NNP'), ('7th', 'CD'), ('July', 'NNP'), ('.', '.'), ('In', 'IN'), ('consonance', 'NN'), ('public', 'JJ'), ('participation', 'NN'), ('.', '.'), ('Maharashtra', 'NNP'), ('Forest', 'NNP'), ('Department', 'NNP'), ('initiated', 'VBD'), ('.', '.'), ('Maharashtra', 'NNP'), ('Harit', 'NNP'), ('Sena', 'NNP'), ('.', '.'), ('/', 'NNP'), ('.', '.'), ('Green', 'NNP'), ('Army', 'NNP'), ('.', '.'), ('body', 'NN'), ('dedicated', 'VBD'), ('volunteers', 'NNS'), ('participate', 'JJ'), ('plantation', 'NN'), ('.', '.'), ('protection', 'NN'), ('.', '.'), ('activities', 'NNS'), ('forest', 'VBP'), ('.', '.'), ('wildlife', 'NN'), ('.', '.'), ('related', 'JJ'), ('sectors', 'NNS'), ('around', 'IN'), ('year', 'NN'), ('.', '.'), ('Individuals', 'NNS'), ('organisations', 'NNS'), ('interested', 'JJ'), ('volunteering', 'VBG'), ('register', 'NN'), ('Green', 'NNP'), ('Army', 'NNP'), ('website', 'NN'), ('www.greenarmy.mahaforest.gov.in', 'VBD'), ('An', 'DT'), ('integrated', 'JJ'), ('drive', 'NN'), ('set', 'VBN'), ('place', 'NN'), ('ensure', 'VB'), ('seamless', 'JJ'), ('successful', 'JJ'), ('participation', 'NN'), ('stakeholders', 'NNS'), ('society', 'NN'), ('.', '.'), ('especially', 'RB'), ('public', 'JJ'), ('.', '.'), ('.')] ]

## Text Stemming

It is modifying a word to obtain its variants using different linguistic processes like affixation (addition of affixes). For example, the stem of the word "studying" is "study", to which -ing.

In [ ]:

```
from nltk.stem import PorterStemmer
from nltk.tokenize import word_tokenize
porter=PorterStemmer()
def stemSentence(sentence):
    token_words=word_tokenize(sentence)
    token_words
    stem_sentence=[]
    for word in token_words:
        stem_sentence.append(porter.stem(word))
```



```

        stem_sentence.append(" ")
    return " ".join(stem_sentence)

example = "Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from data in various forms, both structured and unstructured, similar to data mining."
print(example)
print("-----\nStemmed sentence\n-----")
x=stemSentence(example)
print(x)

```

Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from data in various forms, both structured and unstructured, similar to data mining.

-----

Stemmed sentence

-----

data scienc is an interdisciplinari field that use scientif method , process , algorithm and system to extract knowledg and insight from data in variou f orm , both structur and unstructur , similar to data mine .

### Text lemmatization

It is the process of eliminating redundant prefix or suffix of a word and extract the base word (lemma).

```

In [ ]: from nltk.stem import WordNetLemmatizer
wordnet_lemmatizer = WordNetLemmatizer()
lemmatized_text = []
print("{0:20}{1:20}".format("Word", "Lemma"))
for word in filtered_sentence:
    lemmatized_text.append(wordnet_lemmatizer.lemmatize(word, pos="v"))
    print ("{0:20}{1:20}".format(word, wordnet_lemmatizer.lemmatize(word, pos="v")))
# n->nouns a->adjectives v->verbs

```

Word	Lemma
Between	Between
2016	2016
2019	2019
,	,
state	state
forest	forest
department	department
BJP	BJP
government	government
launched	launch
,	,
Green	Green
Maharashtra	Maharashtra
,	,
drive	drive
aim	aim
plant	plant
50	50
crore	crore
trees	tree
across	across
state	state
four-year	four-year
period	period
.	.
In	In
October	October
2019	2019
,	,
government	government
claimed	claim
surpassed	surpass
target	target
planting	plant
33	33
crore	crore
trees	tree
July-September	July-September
2019	2019
.	.
The	The
Indian	Indian
Express	Express
found	find
non-forest	non-forest
agencies	agencies
—	—
gram	gram
panchayats	panchayats
—	—
tasked	task
planting	plant
trees	tree
uploaded	upload
mandatory	mandatory
audio-visual	audio-visual
proof	proof
tree	tree
plantation	plantation
drives	drive
specially	specially
created	create
portal	portal

.	.
In	In
Pune	Pune
Revenue	Revenue
Division	Division
,	,
claimed	claim
gram	gram
panchayats	panchayats
planted	plant
1.7	1.7
crore	crore
saplings	saplings
;	;
however	however
,	,
evidence	evidence
uploaded	upload
87	87
per	per
cent	cent
(	(
1.49	1.49
crore	crore
)	)
saplings	saplings
.	.
Also	Also
,	,
59	59
government	government
agencies	agencies
involved	involve
drive	drive
many	many
38	38
submitted	submit
survival	survival
reports	report
saplings	saplings
.	.
This	This
year	year
,	,
targets	target
set	set
forest	forest
department	department
comparatively	comparatively
modest	modest
.	.
For	For
example	example
,	,
Pune	Pune
Circle	Circle
—	—
comprises	comprise
three	three
divisions	divisions
Pune	Pune
Solapur	Solapur
district	district
—	—

planned	plan
plant	plant
17	17
lakh	lakh
saplings	saplings
forest	forest
land	land
.	.
However	However
,	,
may	may
meet	meet
target	target
due	due
unavailability	unavailability
funds	fund
.	.
Last	Last
year	year
Pune	Pune
Circle	Circle
planted	plant
70	70
lakh	lakh
saplings	saplings
forest	forest
land	land
.	.
In	In
Pune	Pune
Division	Division
—	—
comprises	comprise
six	six
talukas	talukas
namely	namely
Maval	Maval
,	,
Mulshi	Mulshi
,	,
Daund	Daund
,	,
Indapur	Indapur
,	,
Baramati	Baramati
Havveli	Havveli
—	—
preparations	preparations
done	do
plantation	plantation
4	4
lakh	lakh
trees	tree
special	special
emphasis	emphasis
teakwood	teakwood
.	.
The	The
National	National
Forest	Forest
Policy	Policy
aims	aim
emphasizes	emphasize
maintaining	maintain

33	33
%	%
country	country
,	,
geographical	geographical
area	area
forest	forest
green	green
cover	cover
.	.
In	In
view	view
part	part
50	50
crore	crore
plantation	plantation
programme	programme
within	within
Maharashtra	Maharashtra
,	,
Maharashtra	Maharashtra
Forest	Forest
Department	Department
aims	aim
plant	plant
4	4
crore	crore
saplings	saplings
State	State
1st	1st
July	July
7th	7th
,	,
2017	2017
celebrate	celebrate
,	,
Vanmohotsav	Vanmohotsav
,	,
.	.
The	The
plantation	plantation
programme	programme
,	,
announced	announce
2016	2016
aim	aim
planting	plant
2	2
crore	crore
trees	tree
1st	1st
July	July
2016	2016
resounding	resound
success	success
final	final
total	total
reported	report
figure	figure
2.82	2.82
crore	crore
saplings	saplings
planted	plant
single	single

day	day
.	.
To	To
maintain	maintain
consistency	consistency
platform	platform
without	without
affecting	affect
momentum	momentum
,	,
Forest	Forest
Department	Department
set	set
target	target
plantation	plantation
4	4
crore	crore
,	,
13	13
crore	crore
33	33
crore	crore
saplings	saplings
mission	mission
50	50
crore	crore
plantation	plantation
shall	shall
accomplished	accomplish
three	three
consecutive	consecutive
years	years
viz	viz
.	.
2017	2017
,	,
2018	2018
2019	2019
.	.
The	The
4	4
crore	crore
saplings	saplings
year	year
2017	2017
planted	plant
Vanmohotsav	Vanmohotsav
,	,
July	July
1st	1st
July	July
7th	7th
state-wide	state-wide
drive	drive
involvement	involvement
33	33
government	government
departments	departments
along	along
Students	Students
Schools	Schools
Colleges	Colleges
,	,
NSS	NSS

,	,
NCC	NCC
,	,
CSR	CSR
,	,
NGOs	NGOs
,	,
Railways	Railways
,	,
National Highways	National Highways
,	,
Defence	Defence
,	,
NABARD stakeholders Society	NABARD stakeholders Society
.	.
In first kind	In first kind
,	,
24-hour toll free helpline number 1926 called	24-hour toll free helpline number 1926 call
,	,
Hello Forest	Hello Forest
,	,
set provide information regarding plantation	set provide information regard plantation
,	,
protection mass awareness	protection mass awareness
.	.
The Forest Department created mobile application called	The Forest Department create mobile application call
,	,
My Plants	My Plants
,	,
record details plantation numbers	record detail plantation number
,	,
species location computer system Forest	species location computer system Forest

Department	Department
.	.
All	All
volunteers	volunteer
individual	individual
,	,
collective	collective
organizational	organizational
level	level
download	download
use	use
application	application
record	record
tree	tree
plantation	plantation
work	work
application	application
,	,
operational	operational
1st	1st
July	July
7th	7th
July	July
.	.
In	In
consonance	consonance
public	public
participation	participation
,	,
Maharashtra	Maharashtra
Forest	Forest
Department	Department
initiated	initiate
,	,
Maharashtra	Maharashtra
Harit	Harit
Sena	Sena
,	,
/	/
,	,
Green	Green
Army	Army
,	,
body	body
dedicated	dedicate
volunteers	volunteer
participate	participate
plantation	plantation
,	,
protection	protection
,	,
activities	activities
forest	forest
,	,
wildlife	wildlife
,	,
related	relate
sectors	sectors
around	around
year	year
.	.
Individuals	Individuals
organisations	organisations
interested	interest



volunteering	volunteer
register	register
Green	Green
Army	Army
website	website
www.greenarmy.mahaforest.gov.in	www.greenarmy.mahaforest.gov.in
An	An
integrated	integrate
drive	drive
set	set
place	place
ensure	ensure
seamless	seamless
successful	successful
participation	participation
stakeholders	stakeholders
society	society
,	,
especially	especially
public	public
.	.

### Document Frequency

```
In [ ]: words_set = set()
words_set = words_set.union(set(lemmatized_text))
df = dict.fromkeys(words_set,0)
```

```
In [ ]: for word in lemmatized_text:
df[word]= df[word]+1
```

```
In [ ]: df
```

```
Out[ ]: {'An': 1,
        'NGOs': 1,
        'Pune': 5,
        'tree': 7,
        '87': 1,
        ',': 38,
        '2': 1,
        'specially': 1,
        '38': 1,
        'Railways': 1,
        '2017': 3,
        '13': 1,
        'application': 3,
        'Hello': 1,
        'system': 1,
        'six': 1,
        'State': 1,
        'district': 1,
        'use': 1,
        'create': 2,
        'meet': 1,
        'Defence': 1,
        'Indian': 1,
        'cent': 1,
        'July-September': 1,
        'however': 1,
        'Individuals': 1,
        'register': 1,
        'Harit': 1,
        'body': 1,
        'place': 1,
        'emphasis': 1,
        'Department': 5,
        'platform': 1,
        'seamless': 1,
        'Havveli': 1,
        'consonance': 1,
        ';': 1,
        'Students': 1,
        ''': 6,
        'across': 1,
        'CSR': 1,
        'country': 1,
        'unavailability': 1,
        'years': 1,
        'period': 1,
        'preparations': 1,
        'within': 1,
        'area': 1,
        'comparatively': 1,
        'number': 2,
        'programme': 2,
        'audio-visual': 1,
        'single': 1,
        'per': 1,
        'call': 2,
        'Vanmohotsav': 2,
        '4': 4,
        'National': 2,
        'year': 4,
        ''': 7,
        '7th': 3,
        'Circle': 2,
        'special': 1,
```

'participate': 1,  
'organisations': 1,  
'mobile': 1,  
'surpass': 1,  
'fund': 1,  
'accomplish': 1,  
'download': 1,  
'70': 1,  
'resound': 1,  
'panchayats': 2,  
'This': 1,  
'due': 1,  
'momentum': 1,  
'crore': 13,  
'species': 1,  
'saplings': 9,  
'helpline': 1,  
'Army': 2,  
'successful': 1,  
'NABARD': 1,  
'agencies': 2,  
'Mulshi': 1,  
'survival': 1,  
'The': 5,  
'example': 1,  
'-': 6,  
'Solapur': 1,  
'geographical': 1,  
'success': 1,  
'many': 1,  
'emphasize': 1,  
'kind': 1,  
'1926': 1,  
'To': 1,  
'green': 1,  
')': 1,  
'Maval': 1,  
'59': 1,  
'government': 4,  
'Daund': 1,  
'Schools': 1,  
'awareness': 1,  
'work': 1,  
'level': 1,  
'plan': 1,  
'department': 2,  
'aim': 4,  
'without': 1,  
'lakh': 3,  
'state-wide': 1,  
'mission': 1,  
'along': 1,  
'.'': 21,  
'1st': 4,  
'2.82': 1,  
'individual': 1,  
'dedicate': 1,  
'interest': 1,  
'initiate': 1,  
'land': 2,  
'involve': 1,  
'activities': 1,  
'plant': 10,  
'day': 1,

'first': 1,  
'provide': 1,  
'gram': 2,  
'plantation': 10,  
'collective': 1,  
'around': 1,  
'Division': 2,  
'2016': 3,  
'Indapur': 1,  
'Colleges': 1,  
'July': 6,  
'Sena': 1,  
'final': 1,  
'do': 1,  
'departments': 1,  
'operational': 1,  
'relate': 1,  
'Between': 1,  
'proof': 1,  
'Highways': 1,  
'especially': 1,  
'October': 1,  
'upload': 2,  
'33': 4,  
'mandatory': 1,  
'announce': 1,  
'However': 1,  
'affect': 1,  
'shall': 1,  
'drive': 5,  
'report': 2,  
'integrate': 1,  
'cover': 1,  
'consecutive': 1,  
'talukas': 1,  
'Revenue': 1,  
'Maharashtra': 5,  
'evidence': 1,  
'involvement': 1,  
'public': 2,  
'mass': 1,  
'BJP': 1,  
'namely': 1,  
'three': 2,  
'computer': 1,  
'portal': 1,  
'consistency': 1,  
'total': 1,  
'record': 2,  
'comprise': 2,  
'task': 1,  
'claim': 2,  
'17': 1,  
'2019': 4,  
'launch': 1,  
'find': 1,  
'Forest': 7,  
'Policy': 1,  
'Last': 1,  
'submit': 1,  
'%': 1,  
'24-hour': 1,  
'organizational': 1,  
'/' : 1,

```

'set': 4,
'For': 1,
'2018': 1,
'Plants': 1,
'50': 3,
'My': 1,
'view': 1,
'detail': 1,
'non-forest': 1,
'modest': 1,
'Green': 3,
'forest': 6,
'toll': 1,
'society': 1,
'ensure': 1,
'Express': 1,
'may': 1,
'celebrate': 1,
'In': 6,
'NSS': 1,
'four-year': 1,
'NCC': 1,
'regard': 1,
'Baramati': 1,
'figure': 1,
'information': 1,
'wildlife': 1,
'free': 1,
'website': 1,
'www.greenarmy.mahaforest.gov.in': 1,
'All': 1,
'divisions': 1,
'volunteer': 3,
'teakwood': 1,
'location': 1,
'stakeholders': 2,
'Society': 1,
'1.7': 1,
'(': 1,
'sectors': 1,
'participation': 2,
'state': 2,
'part': 1,
'viz': 1,
'Also': 1,
'protection': 2,
'maintain': 2,
'target': 4,
'1.49': 1}

```

### Term Frequency

```

In [ ]: tf = {}
        corpusCount = len(lemmatized_text)
        for word, count in df.items():
            tf[word] = count / corpusCount

```

```

In [ ]: tf

```

```
Out[ ]: {'An': 0.002136752136752137,  
        'NGOs': 0.002136752136752137,  
        'Pune': 0.010683760683760684,  
        'tree': 0.014957264957264958,  
        '87': 0.002136752136752137,  
        ',': 0.0811965811965812,  
        '2': 0.002136752136752137,  
        'specially': 0.002136752136752137,  
        '38': 0.002136752136752137,  
        'Railways': 0.002136752136752137,  
        '2017': 0.00641025641025641,  
        '13': 0.002136752136752137,  
        'application': 0.00641025641025641,  
        'Hello': 0.002136752136752137,  
        'system': 0.002136752136752137,  
        'six': 0.002136752136752137,  
        'State': 0.002136752136752137,  
        'district': 0.002136752136752137,  
        'use': 0.002136752136752137,  
        'create': 0.004273504273504274,  
        'meet': 0.002136752136752137,  
        'Defence': 0.002136752136752137,  
        'Indian': 0.002136752136752137,  
        'cent': 0.002136752136752137,  
        'July-September': 0.002136752136752137,  
        'however': 0.002136752136752137,  
        'Individuals': 0.002136752136752137,  
        'register': 0.002136752136752137,  
        'Harit': 0.002136752136752137,  
        'body': 0.002136752136752137,  
        'place': 0.002136752136752137,  
        'emphasis': 0.002136752136752137,  
        'Department': 0.010683760683760684,  
        'platform': 0.002136752136752137,  
        'seamless': 0.002136752136752137,  
        'Havveli': 0.002136752136752137,  
        'consonance': 0.002136752136752137,  
        ';': 0.002136752136752137,  
        'Students': 0.002136752136752137,  
        ''': 0.01282051282051282,  
        'across': 0.002136752136752137,  
        'CSR': 0.002136752136752137,  
        'country': 0.002136752136752137,  
        'unavailability': 0.002136752136752137,  
        'years': 0.002136752136752137,  
        'period': 0.002136752136752137,  
        'preparations': 0.002136752136752137,  
        'within': 0.002136752136752137,  
        'area': 0.002136752136752137,  
        'comparatively': 0.002136752136752137,  
        'number': 0.004273504273504274,  
        'programme': 0.004273504273504274,  
        'audio-visual': 0.002136752136752137,  
        'single': 0.002136752136752137,  
        'per': 0.002136752136752137,  
        'call': 0.004273504273504274,  
        'Vanmohotsav': 0.004273504273504274,  
        '4': 0.008547008547008548,  
        'National': 0.004273504273504274,  
        'year': 0.008547008547008548,  
        ''': 0.014957264957264958,  
        '7th': 0.00641025641025641,  
        'Circle': 0.004273504273504274,  
        'special': 0.002136752136752137,
```

'participate': 0.002136752136752137,  
'organisations': 0.002136752136752137,  
'mobile': 0.002136752136752137,  
'surpass': 0.002136752136752137,  
'fund': 0.002136752136752137,  
'accomplish': 0.002136752136752137,  
'download': 0.002136752136752137,  
'70': 0.002136752136752137,  
'resound': 0.002136752136752137,  
'panchayats': 0.004273504273504274,  
'This': 0.002136752136752137,  
'due': 0.002136752136752137,  
'momentum': 0.002136752136752137,  
'crore': 0.027777777777777776,  
'species': 0.002136752136752137,  
'saplings': 0.019230769230769232,  
'helpline': 0.002136752136752137,  
'Army': 0.004273504273504274,  
'successful': 0.002136752136752137,  
'NABARD': 0.002136752136752137,  
'agencies': 0.004273504273504274,  
'Mulshi': 0.002136752136752137,  
'survival': 0.002136752136752137,  
'The': 0.010683760683760684,  
'example': 0.002136752136752137,  
'-': 0.01282051282051282,  
'Solapur': 0.002136752136752137,  
'geographical': 0.002136752136752137,  
'success': 0.002136752136752137,  
'many': 0.002136752136752137,  
'emphasize': 0.002136752136752137,  
'kind': 0.002136752136752137,  
'1926': 0.002136752136752137,  
'To': 0.002136752136752137,  
'green': 0.002136752136752137,  
)': 0.002136752136752137,  
'Maval': 0.002136752136752137,  
'59': 0.002136752136752137,  
'government': 0.008547008547008548,  
'Daund': 0.002136752136752137,  
'Schools': 0.002136752136752137,  
'awareness': 0.002136752136752137,  
'work': 0.002136752136752137,  
'level': 0.002136752136752137,  
'plan': 0.002136752136752137,  
'department': 0.004273504273504274,  
'aim': 0.008547008547008548,  
'without': 0.002136752136752137,  
'lakh': 0.00641025641025641,  
'state-wide': 0.002136752136752137,  
'mission': 0.002136752136752137,  
'along': 0.002136752136752137,  
'.'': 0.04487179487179487,  
'1st': 0.008547008547008548,  
'2.82': 0.002136752136752137,  
'individual': 0.002136752136752137,  
'dedicate': 0.002136752136752137,  
'interest': 0.002136752136752137,  
'initiate': 0.002136752136752137,  
'land': 0.004273504273504274,  
'involve': 0.002136752136752137,  
'activities': 0.002136752136752137,  
'plant': 0.021367521367521368,  
'day': 0.002136752136752137,

'first': 0.002136752136752137,  
'provide': 0.002136752136752137,  
'gram': 0.004273504273504274,  
'plantation': 0.021367521367521368,  
'collective': 0.002136752136752137,  
'around': 0.002136752136752137,  
'Division': 0.004273504273504274,  
'2016': 0.00641025641025641,  
'Indapur': 0.002136752136752137,  
'Colleges': 0.002136752136752137,  
'July': 0.01282051282051282,  
'Sena': 0.002136752136752137,  
'final': 0.002136752136752137,  
'do': 0.002136752136752137,  
'departments': 0.002136752136752137,  
'operational': 0.002136752136752137,  
'relate': 0.002136752136752137,  
'Between': 0.002136752136752137,  
'proof': 0.002136752136752137,  
'Highways': 0.002136752136752137,  
'especially': 0.002136752136752137,  
'October': 0.002136752136752137,  
'upload': 0.004273504273504274,  
'33': 0.008547008547008548,  
'mandatory': 0.002136752136752137,  
'announce': 0.002136752136752137,  
'However': 0.002136752136752137,  
'affect': 0.002136752136752137,  
'shall': 0.002136752136752137,  
'drive': 0.010683760683760684,  
'report': 0.004273504273504274,  
'integrate': 0.002136752136752137,  
'cover': 0.002136752136752137,  
'consecutive': 0.002136752136752137,  
'talukas': 0.002136752136752137,  
'Revenue': 0.002136752136752137,  
'Maharashtra': 0.010683760683760684,  
'evidence': 0.002136752136752137,  
'involvement': 0.002136752136752137,  
'public': 0.004273504273504274,  
'mass': 0.002136752136752137,  
'BJP': 0.002136752136752137,  
'namely': 0.002136752136752137,  
'three': 0.004273504273504274,  
'computer': 0.002136752136752137,  
'portal': 0.002136752136752137,  
'consistency': 0.002136752136752137,  
'total': 0.002136752136752137,  
'record': 0.004273504273504274,  
'comprise': 0.004273504273504274,  
'task': 0.002136752136752137,  
'claim': 0.004273504273504274,  
'17': 0.002136752136752137,  
'2019': 0.008547008547008548,  
'launch': 0.002136752136752137,  
'find': 0.002136752136752137,  
'Forest': 0.014957264957264958,  
'Policy': 0.002136752136752137,  
'Last': 0.002136752136752137,  
'submit': 0.002136752136752137,  
'%': 0.002136752136752137,  
'24-hour': 0.002136752136752137,  
'organizational': 0.002136752136752137,  
'/' : 0.002136752136752137,



```

'set': 0.008547008547008548,
'For': 0.002136752136752137,
'2018': 0.002136752136752137,
'Plants': 0.002136752136752137,
'50': 0.00641025641025641,
'My': 0.002136752136752137,
'view': 0.002136752136752137,
'detail': 0.002136752136752137,
'non-forest': 0.002136752136752137,
'modest': 0.002136752136752137,
'Green': 0.00641025641025641,
'forest': 0.01282051282051282,
'toll': 0.002136752136752137,
'society': 0.002136752136752137,
'ensure': 0.002136752136752137,
'Express': 0.002136752136752137,
'may': 0.002136752136752137,
'celebrate': 0.002136752136752137,
'In': 0.01282051282051282,
'NSS': 0.002136752136752137,
'four-year': 0.002136752136752137,
'NCC': 0.002136752136752137,
'regard': 0.002136752136752137,
'Baramati': 0.002136752136752137,
'figure': 0.002136752136752137,
'information': 0.002136752136752137,
'wildlife': 0.002136752136752137,
'free': 0.002136752136752137,
'website': 0.002136752136752137,
'www.greenarmy.mahaforest.gov.in': 0.002136752136752137,
'All': 0.002136752136752137,
'divisions': 0.002136752136752137,
'volunteer': 0.00641025641025641,
'teakwood': 0.002136752136752137,
'location': 0.002136752136752137,
'stakeholders': 0.004273504273504274,
'Society': 0.002136752136752137,
'1.7': 0.002136752136752137,
'(': 0.002136752136752137,
'sectors': 0.002136752136752137,
'participation': 0.004273504273504274,
'state': 0.004273504273504274,
'part': 0.002136752136752137,
'viz': 0.002136752136752137,
'Also': 0.002136752136752137,
'protection': 0.004273504273504274,
'maintain': 0.004273504273504274,
'target': 0.008547008547008548,
'1.49': 0.002136752136752137}

```

### Inverse Document Frequency

```
In [ ]: import math
```

```
In [ ]: idfDict = {}
N = len(df)
idfDict = df;
for word, val in idfDict.items():
    idfDict[word] = math.log10(N / val)
```

```
In [ ]: idfDict
```



```
Out[ ]: {'An': 2.3820170425748683,
'NGOs': 2.3820170425748683,
'Pune': 1.6830470382388496,
'tree': 1.5369190025606116,
'87': 2.3820170425748683,
',': 0.8022334459580582,
'2': 2.3820170425748683,
'specially': 2.3820170425748683,
'38': 2.3820170425748683,
'Railways': 2.3820170425748683,
'2017': 1.904895787855206,
'13': 2.3820170425748683,
'application': 1.904895787855206,
'Hello': 2.3820170425748683,
'system': 2.3820170425748683,
'six': 2.3820170425748683,
'State': 2.3820170425748683,
'district': 2.3820170425748683,
'use': 2.3820170425748683,
'create': 2.080987046910887,
'meet': 2.3820170425748683,
'Defence': 2.3820170425748683,
'Indian': 2.3820170425748683,
'cent': 2.3820170425748683,
'July-September': 2.3820170425748683,
'however': 2.3820170425748683,
'Individuals': 2.3820170425748683,
'register': 2.3820170425748683,
'Harit': 2.3820170425748683,
'body': 2.3820170425748683,
'place': 2.3820170425748683,
'emphasis': 2.3820170425748683,
'Department': 1.6830470382388496,
'platform': 2.3820170425748683,
'seamless': 2.3820170425748683,
'Havveli': 2.3820170425748683,
'consonance': 2.3820170425748683,
';': 2.3820170425748683,
'Students': 2.3820170425748683,
''': 1.6038657921912247,
'across': 2.3820170425748683,
'CSR': 2.3820170425748683,
'country': 2.3820170425748683,
'unavailability': 2.3820170425748683,
'years': 2.3820170425748683,
'period': 2.3820170425748683,
'preparations': 2.3820170425748683,
'within': 2.3820170425748683,
'area': 2.3820170425748683,
'comparatively': 2.3820170425748683,
'number': 2.080987046910887,
'programme': 2.080987046910887,
'audio-visual': 2.3820170425748683,
'single': 2.3820170425748683,
'per': 2.3820170425748683,
'call': 2.080987046910887,
'Vanmohotsav': 2.080987046910887,
'4': 1.779957051246906,
'National': 2.080987046910887,
'year': 1.779957051246906,
''': 1.5369190025606116,
'7th': 1.904895787855206,
'Circle': 2.080987046910887,
'special': 2.3820170425748683,
```

'participate': 2.3820170425748683,  
'organisations': 2.3820170425748683,  
'mobile': 2.3820170425748683,  
'surpass': 2.3820170425748683,  
'fund': 2.3820170425748683,  
'accomplish': 2.3820170425748683,  
'download': 2.3820170425748683,  
'70': 2.3820170425748683,  
'resound': 2.3820170425748683,  
'panchayats': 2.080987046910887,  
'This': 2.3820170425748683,  
'due': 2.3820170425748683,  
'momentum': 2.3820170425748683,  
'crore': 1.2680736902680316,  
'species': 2.3820170425748683,  
'saplings': 1.4277745331355436,  
'helpline': 2.3820170425748683,  
'Army': 2.080987046910887,  
'successful': 2.3820170425748683,  
'NABARD': 2.3820170425748683,  
'agencies': 2.080987046910887,  
'Mulshi': 2.3820170425748683,  
'survival': 2.3820170425748683,  
'The': 1.6830470382388496,  
'example': 2.3820170425748683,  
'—': 1.6038657921912247,  
'Solapur': 2.3820170425748683,  
'geographical': 2.3820170425748683,  
'success': 2.3820170425748683,  
'many': 2.3820170425748683,  
'emphasize': 2.3820170425748683,  
'kind': 2.3820170425748683,  
'1926': 2.3820170425748683,  
'To': 2.3820170425748683,  
'green': 2.3820170425748683,  
')': 2.3820170425748683,  
'Maval': 2.3820170425748683,  
'59': 2.3820170425748683,  
'government': 1.779957051246906,  
'Daund': 2.3820170425748683,  
'Schools': 2.3820170425748683,  
'awareness': 2.3820170425748683,  
'work': 2.3820170425748683,  
'level': 2.3820170425748683,  
'plan': 2.3820170425748683,  
'department': 2.080987046910887,  
'aim': 1.779957051246906,  
'without': 2.3820170425748683,  
'lakh': 1.904895787855206,  
'state-wide': 2.3820170425748683,  
'mission': 2.3820170425748683,  
'along': 2.3820170425748683,  
'.': 1.059797747840949,  
'1st': 1.779957051246906,  
'2.82': 2.3820170425748683,  
'individual': 2.3820170425748683,  
'dedicate': 2.3820170425748683,  
'interest': 2.3820170425748683,  
'initiate': 2.3820170425748683,  
'land': 2.080987046910887,  
'involve': 2.3820170425748683,  
'activities': 2.3820170425748683,  
'plant': 1.3820170425748683,  
'day': 2.3820170425748683,

'first': 2.3820170425748683,  
'provide': 2.3820170425748683,  
'gram': 2.080987046910887,  
'plantation': 1.3820170425748683,  
'collective': 2.3820170425748683,  
'around': 2.3820170425748683,  
'Division': 2.080987046910887,  
'2016': 1.904895787855206,  
'Indapur': 2.3820170425748683,  
'Colleges': 2.3820170425748683,  
'July': 1.6038657921912247,  
'Sena': 2.3820170425748683,  
'final': 2.3820170425748683,  
'do': 2.3820170425748683,  
'departments': 2.3820170425748683,  
'operational': 2.3820170425748683,  
'relate': 2.3820170425748683,  
'Between': 2.3820170425748683,  
'proof': 2.3820170425748683,  
'Highways': 2.3820170425748683,  
'especially': 2.3820170425748683,  
'October': 2.3820170425748683,  
'upload': 2.080987046910887,  
'33': 1.779957051246906,  
'mandatory': 2.3820170425748683,  
'announce': 2.3820170425748683,  
'However': 2.3820170425748683,  
'affect': 2.3820170425748683,  
'shall': 2.3820170425748683,  
'drive': 1.6830470382388496,  
'report': 2.080987046910887,  
'integrate': 2.3820170425748683,  
'cover': 2.3820170425748683,  
'consecutive': 2.3820170425748683,  
'talukas': 2.3820170425748683,  
'Revenue': 2.3820170425748683,  
'Maharashtra': 1.6830470382388496,  
'evidence': 2.3820170425748683,  
'involvement': 2.3820170425748683,  
'public': 2.080987046910887,  
'mass': 2.3820170425748683,  
'BJP': 2.3820170425748683,  
'namely': 2.3820170425748683,  
'three': 2.080987046910887,  
'computer': 2.3820170425748683,  
'portal': 2.3820170425748683,  
'consistency': 2.3820170425748683,  
'total': 2.3820170425748683,  
'record': 2.080987046910887,  
'comprise': 2.080987046910887,  
'task': 2.3820170425748683,  
'claim': 2.080987046910887,  
'17': 2.3820170425748683,  
'2019': 1.779957051246906,  
'launch': 2.3820170425748683,  
'find': 2.3820170425748683,  
'Forest': 1.5369190025606116,  
'Policy': 2.3820170425748683,  
'Last': 2.3820170425748683,  
'submit': 2.3820170425748683,  
'%': 2.3820170425748683,  
'24-hour': 2.3820170425748683,  
'organizational': 2.3820170425748683,  
'/' : 2.3820170425748683,

```

'set': 1.779957051246906,
'For': 2.3820170425748683,
'2018': 2.3820170425748683,
'Plants': 2.3820170425748683,
'50': 1.904895787855206,
'My': 2.3820170425748683,
'view': 2.3820170425748683,
'detail': 2.3820170425748683,
'non-forest': 2.3820170425748683,
'modest': 2.3820170425748683,
'Green': 1.904895787855206,
'forest': 1.6038657921912247,
'toll': 2.3820170425748683,
'society': 2.3820170425748683,
'ensure': 2.3820170425748683,
'Express': 2.3820170425748683,
'may': 2.3820170425748683,
'celebrate': 2.3820170425748683,
'In': 1.6038657921912247,
'NSS': 2.3820170425748683,
'four-year': 2.3820170425748683,
'NCC': 2.3820170425748683,
'regard': 2.3820170425748683,
'Baramati': 2.3820170425748683,
'figure': 2.3820170425748683,
'information': 2.3820170425748683,
'wildlife': 2.3820170425748683,
'free': 2.3820170425748683,
'website': 2.3820170425748683,
'www.greenarmy.mahaforest.gov.in': 2.3820170425748683,
'All': 2.3820170425748683,
'divisions': 2.3820170425748683,
'volunteer': 1.904895787855206,
'teakwood': 2.3820170425748683,
'location': 2.3820170425748683,
'stakeholders': 2.080987046910887,
'Society': 2.3820170425748683,
'1.7': 2.3820170425748683,
'(': 2.3820170425748683,
'sectors': 2.3820170425748683,
'participation': 2.080987046910887,
'state': 2.080987046910887,
'part': 2.3820170425748683,
'viz': 2.3820170425748683,
'Also': 2.3820170425748683,
'protection': 2.080987046910887,
'maintain': 2.080987046910887,
'target': 1.779957051246906,
'1.49': 2.3820170425748683}

```

TF-IDF

```

In [ ]: tfidf = {}
        for word, val in tf.items():
            tfidf[word] = val * idfDict[word]

```

```

In [ ]: tfidf

```

```
Out[ ]: {'An': 0.005089780005501856,  
        'NGOs': 0.005089780005501856,  
        'Pune': 0.017981271776056088,  
        'tree': 0.022988104739154448,  
        '87': 0.005089780005501856,  
        ',': 0.06513861313334661,  
        '2': 0.005089780005501856,  
        'specially': 0.005089780005501856,  
        '38': 0.005089780005501856,  
        'Railways': 0.005089780005501856,  
        '2017': 0.01221087043496927,  
        '13': 0.005089780005501856,  
        'application': 0.01221087043496927,  
        'Hello': 0.005089780005501856,  
        'system': 0.005089780005501856,  
        'six': 0.005089780005501856,  
        'State': 0.005089780005501856,  
        'district': 0.005089780005501856,  
        'use': 0.005089780005501856,  
        'create': 0.008893107038080715,  
        'meet': 0.005089780005501856,  
        'Defence': 0.005089780005501856,  
        'Indian': 0.005089780005501856,  
        'cent': 0.005089780005501856,  
        'July-September': 0.005089780005501856,  
        'however': 0.005089780005501856,  
        'Individuals': 0.005089780005501856,  
        'register': 0.005089780005501856,  
        'Harit': 0.005089780005501856,  
        'body': 0.005089780005501856,  
        'place': 0.005089780005501856,  
        'emphasis': 0.005089780005501856,  
        'Department': 0.017981271776056088,  
        'platform': 0.005089780005501856,  
        'seamless': 0.005089780005501856,  
        'Havveli': 0.005089780005501856,  
        'consonance': 0.005089780005501856,  
        ';': 0.005089780005501856,  
        'Students': 0.005089780005501856,  
        ''': 0.020562381951169547,  
        'across': 0.005089780005501856,  
        'CSR': 0.005089780005501856,  
        'country': 0.005089780005501856,  
        'unavailability': 0.005089780005501856,  
        'years': 0.005089780005501856,  
        'period': 0.005089780005501856,  
        'preparations': 0.005089780005501856,  
        'within': 0.005089780005501856,  
        'area': 0.005089780005501856,  
        'comparatively': 0.005089780005501856,  
        'number': 0.008893107038080715,  
        'programme': 0.008893107038080715,  
        'audio-visual': 0.005089780005501856,  
        'single': 0.005089780005501856,  
        'per': 0.005089780005501856,  
        'call': 0.008893107038080715,  
        'Vanmohotsav': 0.008893107038080715,  
        '4': 0.015213308130315437,  
        'National': 0.008893107038080715,  
        'year': 0.015213308130315437,  
        ''': 0.022988104739154448,  
        '7th': 0.01221087043496927,  
        'Circle': 0.008893107038080715,  
        'special': 0.005089780005501856,
```

'participate': 0.005089780005501856,  
'organisations': 0.005089780005501856,  
'mobile': 0.005089780005501856,  
'surpass': 0.005089780005501856,  
'fund': 0.005089780005501856,  
'accomplish': 0.005089780005501856,  
'download': 0.005089780005501856,  
'70': 0.005089780005501856,  
'resound': 0.005089780005501856,  
'panchayats': 0.008893107038080715,  
'This': 0.005089780005501856,  
'due': 0.005089780005501856,  
'momentum': 0.005089780005501856,  
'crore': 0.035224269174111984,  
'species': 0.005089780005501856,  
'saplings': 0.02745720256029892,  
'helpline': 0.005089780005501856,  
'Army': 0.008893107038080715,  
'successful': 0.005089780005501856,  
'NABARD': 0.005089780005501856,  
'agencies': 0.008893107038080715,  
'Mulshi': 0.005089780005501856,  
'survival': 0.005089780005501856,  
'The': 0.017981271776056088,  
'example': 0.005089780005501856,  
'-': 0.020562381951169547,  
'Solapur': 0.005089780005501856,  
'geographical': 0.005089780005501856,  
'success': 0.005089780005501856,  
'many': 0.005089780005501856,  
'emphasize': 0.005089780005501856,  
'kind': 0.005089780005501856,  
'1926': 0.005089780005501856,  
'To': 0.005089780005501856,  
'green': 0.005089780005501856,  
')': 0.005089780005501856,  
'Maval': 0.005089780005501856,  
'59': 0.005089780005501856,  
'government': 0.015213308130315437,  
'Daund': 0.005089780005501856,  
'Schools': 0.005089780005501856,  
'awareness': 0.005089780005501856,  
'work': 0.005089780005501856,  
'level': 0.005089780005501856,  
'plan': 0.005089780005501856,  
'department': 0.008893107038080715,  
'aim': 0.015213308130315437,  
'without': 0.005089780005501856,  
'lakh': 0.01221087043496927,  
'state-wide': 0.005089780005501856,  
'mission': 0.005089780005501856,  
'along': 0.005089780005501856,  
'.': 0.047555027146709256,  
'1st': 0.015213308130315437,  
'2.82': 0.005089780005501856,  
'individual': 0.005089780005501856,  
'dedicate': 0.005089780005501856,  
'interest': 0.005089780005501856,  
'initiate': 0.005089780005501856,  
'land': 0.008893107038080715,  
'involve': 0.005089780005501856,  
'activities': 0.005089780005501856,  
'plant': 0.029530278687497188,  
'day': 0.005089780005501856,



'first': 0.005089780005501856,  
'provide': 0.005089780005501856,  
'gram': 0.008893107038080715,  
'plantation': 0.029530278687497188,  
'collective': 0.005089780005501856,  
'around': 0.005089780005501856,  
'Division': 0.008893107038080715,  
'2016': 0.01221087043496927,  
'Indapur': 0.005089780005501856,  
'Colleges': 0.005089780005501856,  
'July': 0.020562381951169547,  
'Sena': 0.005089780005501856,  
'final': 0.005089780005501856,  
'do': 0.005089780005501856,  
'departments': 0.005089780005501856,  
'operational': 0.005089780005501856,  
'relate': 0.005089780005501856,  
'Between': 0.005089780005501856,  
'proof': 0.005089780005501856,  
'Highways': 0.005089780005501856,  
'especially': 0.005089780005501856,  
'October': 0.005089780005501856,  
'upload': 0.008893107038080715,  
'33': 0.015213308130315437,  
'mandatory': 0.005089780005501856,  
'announce': 0.005089780005501856,  
'However': 0.005089780005501856,  
'affect': 0.005089780005501856,  
'shall': 0.005089780005501856,  
'drive': 0.017981271776056088,  
'report': 0.008893107038080715,  
'integrate': 0.005089780005501856,  
'cover': 0.005089780005501856,  
'consecutive': 0.005089780005501856,  
'talukas': 0.005089780005501856,  
'Revenue': 0.005089780005501856,  
'Maharashtra': 0.017981271776056088,  
'evidence': 0.005089780005501856,  
'involvement': 0.005089780005501856,  
'public': 0.008893107038080715,  
'mass': 0.005089780005501856,  
'BJP': 0.005089780005501856,  
'namely': 0.005089780005501856,  
'three': 0.008893107038080715,  
'computer': 0.005089780005501856,  
'portal': 0.005089780005501856,  
'consistency': 0.005089780005501856,  
'total': 0.005089780005501856,  
'record': 0.008893107038080715,  
'comprise': 0.008893107038080715,  
'task': 0.005089780005501856,  
'claim': 0.008893107038080715,  
'17': 0.005089780005501856,  
'2019': 0.015213308130315437,  
'launch': 0.005089780005501856,  
'find': 0.005089780005501856,  
'Forest': 0.022988104739154448,  
'Policy': 0.005089780005501856,  
'Last': 0.005089780005501856,  
'submit': 0.005089780005501856,  
'%': 0.005089780005501856,  
'24-hour': 0.005089780005501856,  
'organizational': 0.005089780005501856,  
'/' : 0.005089780005501856,

```
'set': 0.015213308130315437,  
'For': 0.005089780005501856,  
'2018': 0.005089780005501856,  
'Plants': 0.005089780005501856,  
'50': 0.01221087043496927,  
'My': 0.005089780005501856,  
'view': 0.005089780005501856,  
'detail': 0.005089780005501856,  
'non-forest': 0.005089780005501856,  
'modest': 0.005089780005501856,  
'Green': 0.01221087043496927,  
'forest': 0.020562381951169547,  
'toll': 0.005089780005501856,  
'society': 0.005089780005501856,  
'ensure': 0.005089780005501856,  
'Express': 0.005089780005501856,  
'may': 0.005089780005501856,  
'celebrate': 0.005089780005501856,  
'In': 0.020562381951169547,  
'NSS': 0.005089780005501856,  
'four-year': 0.005089780005501856,  
'NCC': 0.005089780005501856,  
'regard': 0.005089780005501856,  
'Baramati': 0.005089780005501856,  
'figure': 0.005089780005501856,  
'information': 0.005089780005501856,  
'wildlife': 0.005089780005501856,  
'free': 0.005089780005501856,  
'website': 0.005089780005501856,  
'www.greenarmy.mahaforest.gov.in': 0.005089780005501856,  
'All': 0.005089780005501856,  
'divisions': 0.005089780005501856,  
'volunteer': 0.01221087043496927,  
'teakwood': 0.005089780005501856,  
'location': 0.005089780005501856,  
'stakeholders': 0.008893107038080715,  
'Society': 0.005089780005501856,  
'1.7': 0.005089780005501856,  
'(': 0.005089780005501856,  
'sectors': 0.005089780005501856,  
'participation': 0.008893107038080715,  
'state': 0.008893107038080715,  
'part': 0.005089780005501856,  
'viz': 0.005089780005501856,  
'Also': 0.005089780005501856,  
'protection': 0.008893107038080715,  
'maintain': 0.008893107038080715,  
'target': 0.015213308130315437,  
'1.49': 0.005089780005501856}
```