**DSBDAL Assignment no. 7**

import nltk

nltk.download('stopwords')

nltk.download('words')

nltk.download('wordnet')

nltk.download('averged\_perception\_tagger')

nltk.download('punkt')

[nltk\_data] Downloading package stopwords to /root/nltk\_data...

[nltk\_data] Unzipping corpora/stopwords.zip.

[nltk\_data] Downloading package words to /root/nltk\_data...

[nltk\_data] Unzipping corpora/words.zip.

[nltk\_data] Downloading package wordnet to /root/nltk\_data...

[nltk\_data] Error loading averged\_perception\_tagger: Package

[nltk\_data] 'averged\_perception\_tagger' not found in index

[nltk\_data] Downloading package punkt to /root/nltk\_data...

[nltk\_data] Unzipping tokenizers/punkt.zip.

True

import pandas as pd

import numpy as np

sent= ' They told that their ages are 20 23 and 27 respectively'

add=[]

for word in sent.split():

if word.isdigit():

add.append(int(word))

print ('Ave', sum(add)/len(add))

Ave 23.333333333333332

from nltk.tokenize import word\_tokenize, sent\_tokenize

sent= 'Hello all! how are you? Welcome to pune'

sent\_tokenize(sent)

word\_tokenize(sent)

['Hello', 'all', '!', 'how', 'are', 'you', '?', 'Welcome', 'to', 'pune']

from nltk.tokenize import SpaceTokenizer

tk=SpaceTokenizer()

tk.tokenize(sent)

sent='Hello all!\tHow are u?\tto pune'

print(sent)

Hello all! How are u? to pune

s1='ctas','catlike','catty','cat'

s2='stemmer','stemming','stemmed','stem'

s3='fishing','fished','fisher','fish'

s4='argue','argued','argues','argus'

from nltk.stem import PorterStemmer

ps=PorterStemmer()

ps.stem(s3[0])

fish

for word in s4:

ps=PorterStemmer()

print(ps.stem(word))

argu

argu

argu

argu

word='cooking'

from nltk.stem import WordNetLemmatizer

wnl=WordNetLemmatizer()

print(wnl.lemmatize(word,'n')) # noun

print(wnl.lemmatize(word,'v')) # verb

print(wnl.lemmatize(word,'a')) # adjective

print(wnl.lemmatize(word,'r')) # adverb

word='went'

wnl=WordNetLemmatizer()

print(wnl.lemmatize(word,'n')) # noun

cooking

cook

cooking

cooking

went

print(wnl.lemmatize(word,'v')) # verb

print(wnl.lemmatize(word,'a')) # adjective

print(wnl.lemmatize(word,'r')) # adverb

# POS tagging

from nltk import pos\_tag

import nltk

nltk.download('averaged\_perceptron\_tagger')

sents='Rajgad (literal meaning Ruling Fort) is a hill fort situated in the Pune district of Maharashtra, India. Formerly known as Murumdev, the fort was the capital of the Maratha Empire under the rule of Shivaji for almost 26 years, after which the capital was moved to the Raigad Fort.[1] Treasures discovered from an adjacent fort called Torna were used to completely build and fortify the Rajgad Fort.[citation needed]'

print(sents)

go

went

went

[nltk\_data] Downloading package averaged\_perceptron\_tagger to

[nltk\_data] /root/nltk\_data...

[nltk\_data] Unzipping taggers/averaged\_perceptron\_tagger.zip.

Rajgad (literal meaning Ruling Fort) is a hill fort situated in the Pune district of Maharashtra, India. Formerly known as Murumdev, the fort was the capital of the Maratha Empire under the rule of Shivaji for almost 26 years, after which the capital was moved to the Raigad Fort.[1] Treasures discovered from an adjacent fort called Torna were used to completely build and fortify the Rajgad Fort.[citation needed]

words=word\_tokenize(sents)

nltk.download('omw-1.4')

pos\_tag(words)

tags=pos\_tag(words)

for word in tags:

if word[1].startswith('V'):

print(word[0])

words=word\_tokenize(sents)

nltk.download('omw-1.4')

pos\_tag(words)

tags=pos\_tag(words)

for word in tags:

    if word[1].startswith('V'):

        print(word[0])

# spell correction

from textblob import TextBlob

t=TextBlob('computoor')

print(t.correct())

t=TextBlob('nead')

print(t.correct())

computer

head