

LAB ASSIGNMENT - 3

Reg. No.: 21BCE8118

Tagged sentence: [('This', 'DT'), ('is', 'VBZ'), ('an', 'DT'), ('example', 'NN'), ('sentence', None), ('.', '.')]

```
#3. Implement Max Entropy Classifier
from scipy.stats import entropy as en
a=[1,2,3,4]
import pandas as pd
a=pd.Series(a)
data=a.value_counts()
en(data)
```

```
1.3862943611198906
```

```
#4. Implement NER tagger.
import nltk
nltk.download('punkt')
nltk.download('averaged_perceptron_tagger')
nltk.download('maxent_ne_chunker')
nltk.download('words')
from nltk import word_tokenize
from nltk import pos_tag
from nltk import ne_chunk
nltk.download('averaged_perceptron_tagger')
text= 'sundar pichai is one of the CEO of company Google and was born in Chennai.'
tokens=nltk.word_tokenize(text)
print(tokens)
tag=nltk.pos_tag(tokens)
print(tag)
ne=nltk.ne_chunk(tag)
print(ne)
```

```
[('sundar', 'NN'), ('pichai', 'NN'), ('is', 'VBZ'), ('one', 'CD'), ('of', 'IN'), ('the', 'DT'), ('CEO', 'NNP'), ('of', 'IN'), ('company', 'NN'), ('Google', 'NNP'), ('and', 'CC'), ('was', 'VBD'), ('born', 'VBN'), ('in', 'IN'), ('Chennai', 'NNP'), ('.', '.')]
(S
  sundar/NN
  pichai/NN
  is/VBZ
  one/CD
  of/IN
  the/DT
  (ORGANIZATION CEO/NNP)
  of/IN
  company/NN
  (PERSON Google/NNP)
  and/CC
  was/VBD
  born/VBN
  in/IN
  (GPE Chennai/NNP)
  ./.)
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]   /root/nltk_data...
[nltk_data]   Package averaged_perceptron_tagger is already up-to-
[nltk_data]   date!
[nltk_data] Downloading package maxent_ne_chunker to
[nltk_data]   /root/nltk_data...
[nltk_data]   Package maxent_ne_chunker is already up-to-date!
[nltk_data] Downloading package words to /root/nltk_data...
[nltk_data]   Package words is already up-to-date!
```

```
[nltk_data] Error loading averaged_perceptron tagger: Package
[nltk_data]   'averaged_perceptron tagger' not found in index
```

```
# 5.write a tagger that tags Date and Money expressions
```

```
import nltk
```

```
text='The India Cricket team saw a transition after 2011 World Cup'
```

```
result= word_tokenize(text)
```

```
final= nltk.pos_tag(result)
```

```
print(final)
```

```
[('The', 'DT'), ('India', 'NNP'), ('Cricket', 'NNP'), ('team', 'NN'), ('saw', 'VBD'), ('a', 'DT'), ('transition', 'NN'), ('after', 'IN'), ('2011', 'CD'), ('World', 'NNP'), ('Cup', 'NN
```