

In [1]:

```
1 import pandas as pd
2 from nltk import word_tokenize
3 from nltk.tag import CRFTagger
```

In [2]:

```
1 import nltk
2 nltk.download('punkt')
```

[nltk_data] Downloading package punkt to
[nltk_data] C:\Users\USER\AppData\Roaming\nltk_data...
[nltk_data] Package punkt is already up-to-date!

Out[2]:

True

In [3]:

```
1 data=pd.read_excel("hasilpreprocessing.xlsx")
2 data
```

Out[3]:

	Nama	Waktu	Rating	Review	CleanReview
0	Sonia Soraya	3 minggu lalu	5 bintang	Bismillah tadi kita ke ke sana.. Pas masuk di...	bismilah tadi sana pas masuk suru cuci tangan ...
1	Andy Kristyagita	sebulan lalu	5 bintang	Walaupun sudah berulang kali ke sini semenjak ...	walaupun ulang kali sini semenjak kecil rasa b...
2	Edza Tamara	2 minggu lalu	5 bintang	Suasanan malam hari di lokawisata baturraden n...	suasanan malam hari lokawisata baturaden nih s...
3	Abrakadabra	seminggu lalu	5 bintang	Tempat nya sangat Bersih Dan rapih, butuh stam...	tempat sangat bersih rapih butuh stamina utama...
4	Riko Tampati	sebulan lalu	4 bintang	Masih masa lebaran, tiket masuk agak mahal 25r...	masa lebaran tiket masuk mahal rb sangat ramai...
...
285	bunda popon	3 bulan lalu	5 bintang	Paket komplit, murah meriah cocok untuk acara ...	paket komplit murah riah cocok acara keluarga
286	petruk	setahun lalu	5 bintang	Lokasi nya nyaman dan aman, banyak tempat untu...	lokasi nyaman aman banyak tempat berswapphoto b...
287	Eva Meliana	2 tahun lalu	5 bintang	Tempat nya bagus, sejuk, pemandangan nya air nya...	tempat bagus sejuk mandi air dingin hangat fla...
288	Deandra Almerraff W.T	sebulan lalu	5 bintang	View bagus,,, tiket murah, berenang free	view bagus tiket murah renang fre
289	agniya ridha	2 tahun lalu	5 bintang	Seneng banget bisa kesini, suasananya alam ban...	neng banget kesini suasana alam banget bukit s...

290 rows × 5 columns

In [4]:

```
1 review=data['CleanReview']
2 review
```

Out[4]:

0 bismilah tadi sana pas masuk suru cuci tangan ...
1 walaupun ulang kali sini semenjak kecil rasa b...
2 suasanan malam hari lokawisata baturaden nih s...
3 tempat sangat bersih rapih butuh stamina utama...
4 masa lebaran tiket masuk mahal rb sangat ramai...
...
285 paket komplit murah riah cocok acara keluarga
286 lokasi nyaman aman banyak tempat berswapphoto b...
287 tempat bagus sejuk mandi air dingin hangat fla...
288 view bagus tiket murah renang fre
289 neng banget kesini suasana alam banget bukit s...
Name: CleanReview, Length: 290, dtype: object

Tokenizing

In [5]:

```
1 token_word =review.apply(word_tokenize)
2 token_word
```

Out[5]:

```
0      [bismillah, tadi, sana, pas, masuk, suru, cuci,...
1      [walaupun, ulang, kali, sini, semenjak, kecil,...
2      [suasanan, malam, hari, lokawisata, baturaden,...
3      [tempat, sangat, bersih, rapih, butuh, stamina...
4      [masa, lebaran, tiket, masuk, mahal, rb, sanga...
...
285     [paket, komplit, murah, riah, cocok, acara, ke...
286     [lokasi, nyaman, aman, banyak, tempat, berswap...
287     [tempat, bagus, sejuk, mandi, air, dingin, han...
288         [view, bagus, tiket, murah, renang, fre]
289     [neng, banget, kesini, suasana, alam, banget, ...
Name: CleanReview, Length: 290, dtype: object
```

In [6]:

```
1 pip install sklearn-pycrfsuite
```

```
Collecting sklearn-pycrfsuite
  Using cached sklearn_pycrfsuite-0.4.0-py2.py3-none-any.whl
Requirement already satisfied: six in c:\users\user\anaconda3\lib\site-packages (from sklearn-pycrfsuite) (1.16.0)
Collecting python-crfsuite-extension
  Using cached python-crfsuite-extension-0.9.7.tar.gz (485 kB)
Requirement already satisfied: tqdm>=2.0 in c:\users\user\anaconda3\lib\site-packages (from sklearn-pycrfsuite) (4.62.3)
Requirement already satisfied: tabulate in c:\users\user\anaconda3\lib\site-packages (from sklearn-pycrfsuite) (0.9.0)
Requirement already satisfied: colorama in c:\users\user\anaconda3\lib\site-packages (from tqdm>=2.0->sklearn-pycrfsuite) (0.4.4)
Building wheels for collected packages: python-crfsuite-extension
  Building wheel for python-crfsuite-extension (setup.py): started
  Building wheel for python-crfsuite-extension (setup.py): finished with status 'error'
  Running setup.py clean for python-crfsuite-extension
Failed to build python-crfsuite-extension
Installing collected packages: python-crfsuite-extension, sklearn-pycrfsuite
  Running setup.py install for python-crfsuite-extension: started
  Running setup.py install for python-crfsuite-extension: finished with status 'error'
Note: you may need to restart the kernel to use updated packages.
```

```
ERROR: Command errored out with exit status 1:
  command: 'C:\Users\USER\anaconda3\python.exe' -u -c 'import io, os, sys, setuptools, tokenize; sys.argv[0] =
  '''C:\Users\USER\AppData\Local\Temp\pip-install-10a3vqf3\python-crfsuite-extension_469a55f59a9e469caebcf217cbf1ed3\setup.py'''; __file__ = '''C:\Users\USER\AppData\Local\Temp\pip-install-10a3vqf3\python-crfsuite-extension_469a55f59a9e469caebcf217cbf1ed3\setup.py''';f = getattr(tokenize, '''open''', open)(__file__) if os.path.exists(__file__) else io.StringIO(''''from setuptools import setup; setup()''');code = f.read().replace(''\n''', '\n''');f.close();exec(compile(code, __file__, '''exec'''))' bdist_wheel -d 'C:\Users\USER\AppData\Local\Temp\pip-wheel-xc4q8kpj'
```

```
  cwd: C:\Users\USER\AppData\Local\Temp\pip-install-10a3vqf3\python-crfsuite-extension_469a55f59a9e469caebcf217cbf1ed3\
Complete output (12 lines):
  running bdist_wheel
  running build
  running build_py
  creating build
  creating build\lib.win-amd64-3.9
  creating build\lib.win-amd64-3.9\pycrfsuite
  copying pycrfsuite\_dumpparser.py -> build\lib.win-amd64-3.9\pycrfsuite
  copying pycrfsuite\_logparser.py -> build\lib.win-amd64-3.9\pycrfsuite
  copying pycrfsuite\_init__.py -> build\lib.win-amd64-3.9\pycrfsuite
  running build_ext
  building 'pycrfsuite._pycrfsuite' extension
  error: Microsoft Visual C++ 14.0 or greater is required. Get it with "Microsoft C++ Build Tools": https://visualstudio.microsoft.com/visual-cpp-build-tools/ (https://visualstudio.microsoft.com/visual-cpp-build-tools/)
```

```
ERROR: Failed building wheel for python-crfsuite-extension
ERROR: Command errored out with exit status 1:
  command: 'C:\Users\USER\anaconda3\python.exe' -u -c 'import io, os, sys, setuptools, tokenize; sys.argv[0] =
  '''C:\Users\USER\AppData\Local\Temp\pip-install-10a3vqf3\python-crfsuite-extension_469a55f59a9e469caebcf217cbf1ed3\setup.py'''; __file__ = '''C:\Users\USER\AppData\Local\Temp\pip-install-10a3vqf3\python-crfsuite-extension_469a55f59a9e469caebcf217cbf1ed3\setup.py''';f = getattr(tokenize, '''open''', open)(__file__) if os.path.exists(__file__) else io.StringIO(''''from setuptools import setup; setup()''');code = f.read().replace(''\n''', '\n''');f.close();exec(compile(code, __file__, '''exec'''))' install --record 'C:\Users\USER\AppData\Local\Temp\pip-record-61_txeiu\install-record.txt' --single-version-externally-managed --compile --install-headers 'C:\Users\USER\anaconda3\Include\python-crfsuite-extension'
  cwd: C:\Users\USER\AppData\Local\Temp\pip-install-10a3vqf3\python-crfsuite-extension_469a55f59a9e469caebcf217cbf1ed3\
```

```
Complete output (12 lines):
  running install
  running build
  running build_py
  creating build
  creating build\lib.win-amd64-3.9
  creating build\lib.win-amd64-3.9\pycrfsuite
  copying pycrfsuite\_dumpparser.py -> build\lib.win-amd64-3.9\pycrfsuite
  copying pycrfsuite\_logparser.py -> build\lib.win-amd64-3.9\pycrfsuite
  copying pycrfsuite\_init__.py -> build\lib.win-amd64-3.9\pycrfsuite
  running build_ext
  building 'pycrfsuite._pycrfsuite' extension
  error: Microsoft Visual C++ 14.0 or greater is required. Get it with "Microsoft C++ Build Tools": https://visualstudio.microsoft.com/visual-cpp-build-tools/ (https://visualstudio.microsoft.com/visual-cpp-build-tools/)
```

```
ERROR: Command errored out with exit status 1: 'C:\Users\USER\anaconda3\python.exe' -u -c 'import io, os, sys, setuptools, tokenize; sys.argv[0] = '''C:\Users\USER\AppData\Local\Temp\pip-install-10a3vqf3\python-crfsuite-extension_469a55f59a9e469caebcf217cbf1ed3\setup.py'''; __file__ = '''C:\Users\USER\AppData\Local\Temp\pip-install-10a3vqf3\python-crfsuite-extension_469a55f59a9e469caebcf217cbf1ed3\setup.py''';f = getattr(tokenize, '''open''', open)(__file__) if os.path.exists(__file__) else io.StringIO(''''from setuptools import setup; setup()''');code = f.read().replace(''\n''', '\n''');f.close();exec(compile(code, __file__, '''exec'''))' install --record 'C:\Users\USER\AppData\Local\Temp\pip-record-61_txeiu\install-record.txt' --single-version-externally-managed --compile --install-headers 'C:\Users\USER\anaconda3\Include\python-crfsuite-extension' Check the logs for full command output.
```

In [7]:

```
1 pip install python-crfsuite
```

Requirement already satisfied: python-crfsuite in c:\users\user\anaconda3\lib\site-packages (0.9.8)
Note: you may need to restart the kernel to use updated packages.

In [8]:

```
1 import pycrfsuite
```

POS TAGGING

In [9]:

```
1 tag = CRFTagger()  
2 tag.set_model_file('all_indo_man_tag_corpus_model.crf.tagger')
```

In [10]:

```
1 hasil_POS= tag.tag_sents(token_word)  
2 hasil_POS
```

Out[10]:

```
[(['bismillah', 'MD'),  
 ('tadi', 'VB'),  
 ('sana', 'NN'),  
 ('pas', 'JJ'),  
 ('masuk', 'VB'),  
 ('suru', 'NN'),  
 ('cuci', 'NN'),  
 ('tangan', 'NN'),  
 ('pakai', 'VB'),  
 ('masker', 'NN'),  
 ('tugas', 'NN'),  
 ('banyak', 'CD'),  
 ('protokol', 'NN'),  
 ('sehat', 'JJ'),  
 ('banget', 'NN'),  
 ('deh', 'VB'),  
 ('depan', 'NN'),  
 ('dalem', 'FW')]
```

In [11]:

```
1 #mengubah nested list menjadi flat list  
2 flatList = [element for innerList in hasil_POS for element in innerList]  
3 flatList
```

Out[11]:

```
[('bismillah', 'MD'),  
 ('tadi', 'VB'),  
 ('sana', 'NN'),  
 ('pas', 'JJ'),  
 ('masuk', 'VB'),  
 ('suru', 'NN'),  
 ('cuci', 'NN'),  
 ('tangan', 'NN'),  
 ('pakai', 'VB'),  
 ('masker', 'NN'),  
 ('tugas', 'NN'),  
 ('banyak', 'CD'),  
 ('protokol', 'NN'),  
 ('sehat', 'JJ'),  
 ('banget', 'NN'),  
 ('deh', 'VB'),  
 ('depan', 'NN'),  
 ('dalem', 'FW')]
```

In [12]:

```

1 import nltk
2 nltk.download('averaged_perceptron_tagger')
3 nltk.download('maxent_ne_chunker')
4 nltk.download('words')

```

```

[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data] C:\Users\USER\AppData\Roaming\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] C:\Users\USER\AppData\Roaming\nltk_data...
[nltk_data] Package maxent_ne_chunker is already up-to-date!
[nltk_data] Downloading package words to
[nltk_data] C:\Users\USER\AppData\Roaming\nltk_data...
[nltk_data] Package words is already up-to-date!

```

Out[12]:

True

NER

In [13]:

```

1 entities =[]
2 labels =[]
3
4 for i,j in flatList:
5     if i and j != None:
6         entities.append(i)
7         labels.append(j)
8
9 entities_labels = list(set(zip(entities, labels)))
10 entities_df = pd.DataFrame(entities_labels)
11 entities_df.columns = ["Entities", "Labels"]
12 entities_df

```

Out[13]:

	Entities	Labels
0	gantung	NN
1	wahana	NN
2	derita	NN
3	kena	SC
4	strategis	JJ
...
1760	buat	JJ
1761	kelelawar	VB
1762	pohon	NN
1763	betah	VB
1764	gengs	NN

1765 rows × 2 columns