In [3]:

```
#20BCE2711
#Task 1
# http://bit.ly/3IoI5JF

def gender_features(word):
    return{'last_letter':word[-1]}
```

In [24]:

```
gender_features('Santhosh')
```

Out[24]:

```
{'last_letter': 'h'}
```

In [6]:

```
from nltk.corpus import names
```

In [10]:

```
labeled_names = ([(name, 'male') for name in names.words('male.txt')] + [(name, 'female') for name in names.words('female.txt')])
```

In [11]:

```
labeled_names
```

```
 ('Andros', 'male'),
 ('Andrus', 'male'),
 ('Andrzej', 'male'),
 ('Andy', 'male'),
 ('Angel', 'male'),
 ('Angelico', 'male'),
 ('Angelo', 'male'),
 ('Angie', 'male'),
 ('Angus', 'male'),
 ('Ansel', 'male'),
 ('Ansell', 'male'),
 ('Anselm', 'male'),
 ('Anson', 'male'),
 ('Anthony', 'male'),
 ('Antin', 'male'),
 ('Antoine', 'male'),
 ('Anton', 'male'),
 ('Antone', 'male'),
 ('Antoni', 'male'),
 ('Antonin', 'male'),
```

In [12]:

```
import random
random.shuffle(labeled_names)
```

In [13]:

```
featuresets = [(gender_features(n), gender) for (n, gender) in labeled_names]
```

In [14]:

```
train_set, test_set = featuresets[500:],featuresets[:500]
```

In [15]:

```
import nltk
classifier = nltk.NaiveBayesClassifier.train(train_set)
```

In [16]:

```
classifier.classify(gender_features('Obama'))
```

Out[16]:

```
'female'
```

In [18]:

```
print(nltk.classify.accuracy(classifier, test_set))
```

```
0.754
```

In [20]:

```
# task 2
import nltk
from nltk.tokenize import TweetTokenizer
```

In [21]:

```
text = 'The party was soo fun :D #superfun'
```

In [22]:

```
twtkn = TweetTokenizer()
```

In [23]:

```
twtkn.tokenize(text)
```

Out[23]:

```
['The', 'party', 'was', 'soo', 'fun', ':D', '#superfun']
```

In [ ]: