J047_LabExam_CLNLP

October 29, 2021

Roshan Srivastava

J047 MBA Tech. EXTC SEM 7

Computational Linguistics and Natural Language Processing Lab Lab Exam

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

```
{'last_letter': 'n'}
```

```python
from nltk.corpus import names
labeled_names = ([(name, 'male') for name in names.words('male.txt')] + [(name,
'female') for name in names.words('female.txt')])
import random
random.shuffle(labeled_names)
```

```python
import nltk
featuresets = [(gender_features(n), gender) for (n, gender) in labeled_names]
train_set, test_set = featuresets[500:], featuresets[:500]
classifier = nltk.NaiveBayesClassifier.train(train_set)
```

```python
classifier.classify(gender_features("Roshan"))
```

```
'male'
```

```python
classifier.classify(gender_features("Chris"))
```

```
'male'
```

```python
classifier.classify(gender_features("Shirley"))
```

```
'female'
```

```python
classifier.show_most_informative_features(5)
```

```
Most Informative Features
             last_letter = 'a'              female : male   =     38.1 : 1.0
             last_letter = 'k'                male : female =     31.2 : 1.0
```

1

```
last_letter = 'f'              male : female =     25.4 : 1.0
last_letter = 'p'              male : female =     11.9 : 1.0
last_letter = 'v'              male : female =     11.2 : 1.0
```