

exp7

October 31, 2021

Neha Kulkarni

A060 BTech. IT SEM 7

Computational Linguistics and Natural Language Processing Lab Experiment 7

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

[1]: {'last_letter': 'k'}
```

```
[2]: 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)
```

```
[3]: 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)

[4]: classifier.classify(gender_features('Neo'))
```

```
[4]: 'male'
```

```
[5]: classifier.classify(gender_features('Trinity'))
```

```
[5]: 'female'
```

```
[6]: print(nltk.classify.accuracy(classifier, test_set))
```

```
0.75
```

```
[7]: classifier.show_most_informative_features(5)
```

Most Informative Features

last_letter = 'a'	female : male =	34.5 : 1.0
last_letter = 'k'	male : female =	31.8 : 1.0
last_letter = 'f'	male : female =	16.0 : 1.0

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

```
[8]: from nltk.classify import apply_features
train_set = apply_features(gender_features, labeled_names[500:])
test_set = apply_features(gender_features, labeled_names[:500])
```

```
[9]: def gender_features2(name):
    features = {}
    features["first_letter"] = name[0].lower()
    features["last_letter"] = name[-1].lower()
    for letter in 'abcdefghijklmnopqrstuvwxyz':
        features["count({})".format(letter)] = name.lower().count(letter)
        features["has({})".format(letter)] = (letter in name.lower())
    return features
```

```
[10]: gender_features2('John')
```

```
[10]: {'first_letter': 'j',
       'last_letter': 'n',
       'count(a)': 0,
       'has(a)': False,
       'count(b)': 0,
       'has(b)': False,
       'count(c)': 0,
       'has(c)': False,
       'count(d)': 0,
       'has(d)': False,
       'count(e)': 0,
       'has(e)': False,
       'count(f)': 0,
       'has(f)': False,
       'count(g)': 0,
       'has(g)': False,
       'count(h)': 1,
       'has(h)': True,
       'count(i)': 0,
       'has(i)': False,
       'count(j)': 1,
       'has(j)': True,
       'count(k)': 0,
       'has(k)': False,
       'count(l)': 0,
       'has(l)': False,
       'count(m)': 0,
       'has(m)': False,
       'count(n)': 1,
       'has(n)': True,
```

```
'count(o)': 1,
'has(o)': True,
'count(p)': 0,
'has(p)': False,
'count(q)': 0,
'has(q)': False,
'count(r)': 0,
'has(r)': False,
'count(s)': 0,
'has(s)': False,
'count(t)': 0,
'has(t)': False,
'count(u)': 0,
'has(u)': False,
'count(v)': 0,
'has(v)': False,
'count(w)': 0,
'has(w)': False,
'count(x)': 0,
'has(x)': False,
'count(y)': 0,
'has(y)': False,
'count(z)': 0,
'has(z)': False}
```

```
[11]: featuresets = [(gender_features2(n), gender) for (n, gender) in labeled_names]
train_set, test_set = featuresets[500:], featuresets[:500]
classifier = nltk.NaiveBayesClassifier.train(train_set)
print(nltk.classify.accuracy(classifier, test_set))
```

0.784

```
[12]: train_names = labeled_names[1500:]
devtest_names = labeled_names[500:1500]
test_names = labeled_names[:500]
```

```
[14]: train_set = [(gender_features(n), gender) for (n, gender) in train_names]
devtest_set = [(gender_features(n), gender) for (n, gender) in devtest_names]
test_set = [(gender_features(n), gender) for (n, gender) in test_names]
classifier = nltk.NaiveBayesClassifier.train(train_set)
print(nltk.classify.accuracy(classifier, devtest_set))
```

0.774

```
[15]: errors = []
for (name, tag) in devtest_names:
    guess = classifier.classify(gender_features(name))
    if guess != tag:
```

```
errors.append((tag, guess, name))

[16]: for tag, guess, name in sorted(errors):
        print('correct={:<8} guess={:<8s} name={:<30}'.format(tag, guess, name))
```

correct=female	guess=ma	name=Abigail
correct=female	guess=ma	name=Adrian
correct=female	guess=ma	name=Alison
correct=female	guess=ma	name=April
correct=female	guess=ma	name=Arabel
correct=female	guess=ma	name=Ayn
correct=female	guess=ma	name=Bert
correct=female	guess=ma	name=Bridget
correct=female	guess=ma	name=Carmel
correct=female	guess=ma	name=Caron
correct=female	guess=ma	name=Caryn
correct=female	guess=ma	name=Cass
correct=female	guess=ma	name=Catlin
correct=female	guess=ma	name=Charis
correct=female	guess=ma	name=Christan
correct=female	guess=ma	name=Christian
correct=female	guess=ma	name=Chrystel
correct=female	guess=ma	name=Cindelyn
correct=female	guess=ma	name=Clair
correct=female	guess=ma	name=Cristin
correct=female	guess=ma	name=Darryl
correct=female	guess=ma	name=Daveen
correct=female	guess=ma	name=Dawn
correct=female	guess=ma	name=Denys
correct=female	guess=ma	name=Devan
correct=female	guess=ma	name=Devon
correct=female	guess=ma	name=Diann
correct=female	guess=ma	name=Doralin
correct=female	guess=ma	name=Ealasaид
correct=female	guess=ma	name=Eilis
correct=female	guess=ma	name=Em
correct=female	guess=ma	name=Enid
correct=female	guess=ma	name=Estell
correct=female	guess=ma	name=Ethelind
correct=female	guess=ma	name=Farrand
correct=female	guess=ma	name=Francis
correct=female	guess=ma	name=Gaynor
correct=female	guess=ma	name=Gill
correct=female	guess=ma	name=Gretal
correct=female	guess=ma	name=Ines
correct=female	guess=ma	name=Isador
correct=female	guess=ma	name=Ivett
correct=female	guess=ma	name=Janot

correct=female	guess=male	name=Jean
correct=female	guess=male	name=Jessalin
correct=female	guess=male	name=Jo
correct=female	guess=male	name=Joell
correct=female	guess=male	name=Joellen
correct=female	guess=male	name=Jonell
correct=female	guess=male	name=Jordain
correct=female	guess=male	name=Joselyn
correct=female	guess=male	name=Joyan
correct=female	guess=male	name=Joyann
correct=female	guess=male	name=Karlen
correct=female	guess=male	name=Kass
correct=female	guess=male	name=Katharyn
correct=female	guess=male	name=Keriann
correct=female	guess=male	name=Kristin
correct=female	guess=male	name=Krystal
correct=female	guess=male	name=Kylen
correct=female	guess=male	name=Lark
correct=female	guess=male	name=Laurel
correct=female	guess=male	name=Lauren
correct=female	guess=male	name=Lauryn
correct=female	guess=male	name=Leland
correct=female	guess=male	name=Lillis
correct=female	guess=male	name=Lorilyn
correct=female	guess=male	name=Lorrin
correct=female	guess=male	name=Mab
correct=female	guess=male	name=Mabel
correct=female	guess=male	name=Madlin
correct=female	guess=male	name=Maren
correct=female	guess=male	name=Marie-Ann
correct=female	guess=male	name=Maryann
correct=female	guess=male	name=Marylou
correct=female	guess=male	name=Meg
correct=female	guess=male	name=Meghann
correct=female	guess=male	name=Merilyn
correct=female	guess=male	name=Milissent
correct=female	guess=male	name=Miran
correct=female	guess=male	name=Mureil
correct=female	guess=male	name=Peg
correct=female	guess=male	name=Pegeen
correct=female	guess=male	name=Robin
correct=female	guess=male	name=Rosamond
correct=female	guess=male	name=Roslyn
correct=female	guess=male	name=Scarlet
correct=female	guess=male	name=Shannon
correct=female	guess=male	name=Shanon
correct=female	guess=male	name=Sherill
correct=female	guess=male	name=Shir

correct=female	guess=male	name=Shirl
correct=female	guess=male	name=Shirleen
correct=female	guess=male	name=Tamiko
correct=female	guess=male	name=Thomasin
correct=female	guess=male	name=Wallis
correct=male	guess=female	name=Addie
correct=male	guess=female	name=Adolph
correct=male	guess=female	name=Amory
correct=male	guess=female	name=Andrey
correct=male	guess=female	name=Andy
correct=male	guess=female	name=Aristotle
correct=male	guess=female	name=Arne
correct=male	guess=female	name=Ashish
correct=male	guess=female	name=Baily
correct=male	guess=female	name=Barnabe
correct=male	guess=female	name=Beale
correct=male	guess=female	name=Bennie
correct=male	guess=female	name=Bjorne
correct=male	guess=female	name=Broddy
correct=male	guess=female	name=Brooke
correct=male	guess=female	name=Bryce
correct=male	guess=female	name=Carlyle
correct=male	guess=female	name=Cary
correct=male	guess=female	name=Chaddie
correct=male	guess=female	name=Chaunce
correct=male	guess=female	name=Clare
correct=male	guess=female	name=Clemente
correct=male	guess=female	name=Cobby
correct=male	guess=female	name=Constantine
correct=male	guess=female	name=Corrie
correct=male	guess=female	name=Daffy
correct=male	guess=female	name=Danie
correct=male	guess=female	name=Davidde
correct=male	guess=female	name=Davide
correct=male	guess=female	name=Dominique
correct=male	guess=female	name=Dory
correct=male	guess=female	name=Durante
correct=male	guess=female	name=Dwane
correct=male	guess=female	name=Earle
correct=male	guess=female	name=Ferdy
correct=male	guess=female	name=Filipe
correct=male	guess=female	name=Fonzie
correct=male	guess=female	name=Franky
correct=male	guess=female	name=Gabriele
correct=male	guess=female	name=Garcia
correct=male	guess=female	name=Garry
correct=male	guess=female	name=Geoffrey
correct=male	guess=female	name=Geoffry

correct=male	guess=female	name=Georgie
correct=male	guess=female	name=Gere
correct=male	guess=female	name=Gerri
correct=male	guess=female	name=Godfree
correct=male	guess=female	name=Graehme
correct=male	guess=female	name=Guthrie
correct=male	guess=female	name=Hadley
correct=male	guess=female	name=Herculie
correct=male	guess=female	name=Hillary
correct=male	guess=female	name=Hirsch
correct=male	guess=female	name=Holly
correct=male	guess=female	name=Hugh
correct=male	guess=female	name=Ikey
correct=male	guess=female	name=Jaime
correct=male	guess=female	name=Jean-Christophe
correct=male	guess=female	name=Jedediah
correct=male	guess=female	name=Jeramie
correct=male	guess=female	name=Jodie
correct=male	guess=female	name=Joe
correct=male	guess=female	name=Juanita
correct=male	guess=female	name=Judith
correct=male	guess=female	name=Kingsly
correct=male	guess=female	name=Larry
correct=male	guess=female	name=Lesley
correct=male	guess=female	name=Locke
correct=male	guess=female	name=Luce
correct=male	guess=female	name=Mackenzie
correct=male	guess=female	name=Manish
correct=male	guess=female	name=Marietta
correct=male	guess=female	name=Matty
correct=male	guess=female	name=Merry
correct=male	guess=female	name=Micah
correct=male	guess=female	name=Mickey
correct=male	guess=female	name=Mordecai
correct=male	guess=female	name=Myke
correct=male	guess=female	name=Neddie
correct=male	guess=female	name=Nickey
correct=male	guess=female	name=Nikki
correct=male	guess=female	name=Obadiah
correct=male	guess=female	name=Orbadiah
correct=male	guess=female	name=Peirce
correct=male	guess=female	name=Pennie
correct=male	guess=female	name=Prince
correct=male	guess=female	name=Pryce
correct=male	guess=female	name=Raleigh
correct=male	guess=female	name=Ramsay
correct=male	guess=female	name=Ramsey
correct=male	guess=female	name=Randolph

```

correct=male    guess=female    name=Randy
correct=male    guess=female    name=Reilly
correct=male    guess=female    name=Roarke
correct=male    guess=female    name=Rocky
correct=male    guess=female    name=Rodge
correct=male    guess=female    name=Ronnie
correct=male    guess=female    name=Rory
correct=male    guess=female    name=Ruby
correct=male    guess=female    name=Salomone
correct=male    guess=female    name=Sammy
correct=male    guess=female    name=Samuele
correct=male    guess=female    name=Scarface
correct=male    guess=female    name=Seth
correct=male    guess=female    name=Sky
correct=male    guess=female    name=Sloane
correct=male    guess=female    name=Sollie
correct=male    guess=female    name=Stanly
correct=male    guess=female    name=Temple
correct=male    guess=female    name=Terri
correct=male    guess=female    name=Tobie
correct=male    guess=female    name=Toby
correct=male    guess=female    name=Toddie
correct=male    guess=female    name=Toddy
correct=male    guess=female    name=Troy
correct=male    guess=female    name=Tully
correct=male    guess=female    name=Udale
correct=male    guess=female    name=Vasily
correct=male    guess=female    name=Wally
correct=male    guess=female    name=Walsh
correct=male    guess=female    name=Wayne
correct=male    guess=female    name=Welbie
correct=male    guess=female    name=Weslie
correct=male    guess=female    name=Winnie
correct=male    guess=female    name=Witty
correct=male    guess=female    name=Woodie
correct=male    guess=female    name=Zach
correct=male    guess=female    name=Zane
correct=male    guess=female    name=Zebadiah
correct=male    guess=female    name=Zolly

```

```
[17]: def gender_features(word):
        return {'suffix1': word[-1:], 'suffix2': word[-2:]}
```

```
[18]: train_set = [(gender_features(n), gender) for (n, gender) in train_names]
devtest_set = [(gender_features(n), gender) for (n, gender) in devtest_names]
classifier = nltk.NaiveBayesClassifier.train(train_set)
print(nltk.classify.accuracy(classifier, devtest_set))
```

0.788

[]: