

In [1]:

```
#import all the necessary libraries
from nltk.corpus import names
import nltk
import random
```

In [2]:

```
def genderFunc(word):
    return {'lastChar': word[-1]}
```

In [3]:

```
#get the last letter
genderFunc("Obama")
```

Out[3]:

```
{'lastChar': 'a'}
```

In [4]:

```
#import all the names
names.words()
```

Out[4]:

```
['Abagael',
 'Abigail',
 'Abbe',
 'Abbey',
 'Abbi',
 'Abbie',
 'Abby',
 'Abigael',
 'Abigail',
 'Abigale',
 'Abra',
 'Acacia',
 'Ada',
 'Adah',
 'Adaline',
 'Adara',
 'Addie',
 'Addis',
 'Adel',
 'Adela',
 'Adelaide',
 'Adele',
 'Adelice',
 'Adelina',
 'Adelind',
 'Adeline',
 'Adella',
 'Adelle',
 'Adena',
 'Adey',
 'Adi',
 'Adiana',
 'Adina',
 'Adora',
```

'Adore',
'Adoree',
'Adorne',
'Adrea',
'Adria',
'Adriaens',
'Adrian',
'Adriana',
'Adriane',
'Adrianna',
'Adrienne',
'Adrien',
'Adriena',
'Adrienne',
'Aeriel',
'Aeriela',
'Aeriell',
'Ag',
'Agace',
'Agata',
'Agatha',
'Agathe',
'Aggi',
'Aggie',
'Aggy',
'Agn',
'Agnella',
'Agnese',
'Agnese',
'Agnese',
'Agneta',
'Agnola',
'Agretha',
'Aida',
'Aidan',
'Aigneis',
'Aila',
'Aile',
'Ailee',
'Aileen',
'Ailene',
'Ailey',
'Aili',
'Ailina',
'Ailyn',
'Aime',
'Aimee',
'Aimil',
'Aina',
'Aindrea',
'Ainslee',
'Ainsley',
'Ainslie',
'Ajay',
'Alaine',
'Alameda',
'Alana',
'Alanah',
'Alane',
'Alanna',
'Alayne',
'Alberta',
'Albertina',
'Albertine',
'Albina',

'Alecia',
'Aleda',
'Aleece',
'Aleecia',
'Aleen',
'Alejandra',
'Alejandrina',
'Alena',
'Alene',
'Alessandra',
'Aleta',
'Alethea',
'Alex',
'Alexa',
'Alexandra',
'Alexandrina',
'Alexi',
'Alexia',
'Alexina',
'Alexine',
'Alexis',
'Alfie',
'Alfreda',
'Ali',
'Alia',
'Alica',
'Alice',
'Alicea',
'Alicia',
'Alida',
'Alidia',
'Alina',
'Aline',
'Alis',
'Alisa',
'Alisha',
'Alison',
'Alissa',
'Alisun',
'Alix',
'Aliza',
'Alla',
'Alleen',
'Allegra',
'Allene',
'Alli',
'Allianora',
'Allie',
'Allina',
'Allis',
'Allison',
'Allissa',
'Allsun',
'Ally',
'Allyce',
'Allyn',
'Allys',
'Allyson',
'Alma',
'Almeda',
'Almeria',
'Almeta',
'Almira',
'Almire',
'Aloise',

'Aloisia',
'Aloysia',
'Alpa',
'Alta',
'Althea',
'Alvera',
'Alvina',
'Alvinia',
'Alvira',
'Alyce',
'Alyda',
'Alys',
'Alysa',
'Alyse',
'Alysia',
'Alyson',
'Alyss',
'Alyssa',
'Amabel',
'Amabelle',
'Amalea',
'Amalee',
'Amaleta',
'Amalia',
'Amalie',
'Amalita',
'Amalle',
'Amanda',
'Amandi',
'Amandie',
'Amandy',
'Amara',
'Amargo',
'Amata',
'Amber',
'Amberly',
'Ambrosia',
'Ambur',
'Ame',
'Amelia',
'Amelie',
'Amelina',
'Ameline',
'Amelita',
'Ami',
'Amie',
'Amity',
'Ammamaria',
'Amy',
'Ana',
'Anabel',
'Anabella',
'Anabelle',
'Anais',
'Analiese',
'Analise',
'Anallese',
'Anallise',
'Anastasia',
'Anastasie',
'Anastassia',
'Anatola',
'Andee',
'Andi',
'Andie',

'Andra',
'Andrea',
'Andreana',
'Andree',
'Andrei',
'Andria',
'Andriana',
'Andriette',
'Andromache',
'Andromeda',
'Andy',
'Anestassia',
'Anet',
'Anett',
'Anetta',
'Anette',
'Ange',
'Angel',
'Angela',
'Angele',
'Angelina',
'Angelica',
'Angelika',
'Angelina',
'Angeline',
'Angelique',
'Angelita',
'Angelle',
'Angie',
'Angil',
'Angy',
'Ania',
'Anica',
'Anissa',
'Anita',
'Anitra',
'Anja',
'Anjanette',
'Anjela',
'Ann',
'Ann-Mari',
'Ann-Marie',
'Anna',
'Anna-Diana',
'Anna-Diane',
'Anna-Maria',
'Annabal',
'Annabel',
'Annabela',
'Annabell',
'Annabella',
'Annabelle',
'Annadiana',
'Annadiane',
'Annalee',
'Annalena',
'Annaliese',
'Annalisa',
'Annalise',
'Annalyse',
'Annamari',
'Annamaria',
'Annamarie',
'Anne',
'Anne-Corinne',

'Anne-Mar',
'Anne-Marie',
'AnneCorinne',
'Anneliese',
'Annelise',
'Annemarie',
'Annetta',
'Annette',
'Anni',
'Annice',
'Annie',
'Annissa',
'Annmaria',
'Annmarie',
'Annnora',
'Annora',
'Anny',
'Anselma',
'Ansley',
'Anstice',
'Anthe',
'Anthea',
'Anthia',
'Antoinette',
'Antonella',
'Antonetta',
'Antonia',
'Antonie',
'Antonietta',
'Antonina',
'Anya',
'Aphrodite',
'Appolonia',
'April',
'Aprilette',
'Ara',
'Arabel',
'Arabela',
'Arabele',
'Arabella',
'Arabelle',
'Arda',
'Ardath',
'Ardeen',
'Ardelia',
'Ardelis',
'Ardella',
'Ardelle',
'Arden',
'Ardene',
'Ardenia',
'Ardine',
'Ardis',
'Ardith',
'Ardra',
'Ardyce',
'Ardys',
'Ardyth',
'Aretha',
'Ariadne',
'Ariana',
'Arianne',
'Aridatha',
'Ariel',
'Ariela',

'Ariella',
'Arielle',
'Arlana',
'Arlee',
'Arleen',
'Arlen',
'Arlena',
'Arlene',
'Arleta',
'Arlette',
'Arleyne',
'Arlie',
'Arliene',
'Arlina',
'Arlinda',
'Arline',
'Arly',
'Arlyn',
'Arlyne',
'Aryn',
'Ashely',
'Ashlee',
'Ashleigh',
'Ashlen',
'Ashley',
'Ashli',
'Ashlie',
'Ashly',
'Asia',
'Astra',
'Astrid',
'Astrix',
'Atalanta',
'Athena',
'Athene',
'Atlanta',
'Atlante',
'Auberta',
'Aubine',
'Aubree',
'Aubrette',
'Aubrey',
'Aubrie',
'Aubry',
'Audi',
'Audie',
'Audra',
'Audre',
'Audrey',
'Audrie',
'Audry',
'Audrye',
'Audy',
'Augusta',
'Auguste',
'Augustina',
'Augustine',
'Aura',
'Aurea',
'Aurel',
'Aurelea',
'Aurelia',
'Aurelie',
'Auria',
'Aurie',

'Aurilia',
'Aurlie',
'Auroora',
'Aurora',
'Aurore',
'Austin',
'Austina',
'Austine',
'Ava',
'Aveline',
'Averil',
'Averyl',
'Avie',
'Avis',
'Aviva',
'Avivah',
'Avril',
'Avrit',
'Ayn',
'Bab',
'Babara',
'Babette',
'Babita',
'Babs',
'Bambi',
'Bambie',
'Bamby',
'Barb',
'Barbabra',
'Barbara',
'Barbara-Anne',
'Barbaraanne',
'Barbe',
'Barbee',
'Barbette',
'Barbey',
'Barbi',
'Barbie',
'Barbra',
'Barby',
'Bari',
'Barrie',
'Barry',
'Basia',
'Bathsheba',
'Batsheva',
'Bea',
'Beatrice',
'Beatrisa',
'Beatrix',
'Beatriz',
'Beau',
'Bebe',
'Becca',
'Becka',
'Becki',
'Beckie',
'Becky',
'Bee',
'Beilul',
'Beitris',
'Bekki',
'Bel',
'Belia',
'Belicia',

'Belinda',
'Belita',
'Bell',
'Bella',
'Bellamy',
'Bellanca',
'Belle',
'Bellina',
'Belva',
'Belvia',
'Bendite',
'Benedetta',
'Benedicta',
'Benedikta',
'Benetta',
'Benita',
'Benni',
'Bennie',
'Benny',
'Benoite',
'Berenice',
'Beret',
'Berget',
'Berna',
'Bernadene',
'Bernadette',
'Bernadina',
'Bernadine',
'Bernardina',
'Bernardine',
'Bernelle',
'Bernete',
'Bernetta',
'Bernette',
'Berni',
'Bernice',
'Bernie',
'Bernita',
'Berny',
'Berri',
'Berrie',
'Berry',
'Bert',
'Berta',
'Berte',
'Bertha',
'Berthe',
'Berti',
'Bertie',
'Bertina',
'Bertine',
'Berty',
'Beryl',
'Beryle',
'Bess',
'Bessie',
'Bessy',
'Beth',
'Bethanne',
'Bethany',
'Bethena',
'Bethina',
'Betsey',
'Betsy',
'Betta',

'Bette',
'Bette-Ann',
'Betteann',
'Betteanne',
'Betti',
'Bettie',
'Bettina',
'Bettine',
'Betty',
'Bettye',
'Beulah',
'Bev',
'Beverie',
'Beverlee',
'Beverlie',
'Beverly',
'Bevvy',
'Bianca',
'Bianka',
'Biddy',
'Bidget',
'Bill',
'Billi',
'Billie',
'Billy',
'Binni',
'Binnie',
'Binny',
'Bird',
'Birdie',
'Birgit',
'Birgitta',
'Blair',
'Blaire',
'Blake',
'Blakelee',
'Blakeley',
'Blanca',
'Blanch',
'Blancha',
'Blanche',
'Blinni',
'Blinnie',
'Blinny',
'Bliss',
'Blisse',
'Blithe',
'Blondell',
'Blondelle',
'Blondie',
'Blondy',
'Blythe',
'Bo',
'Bobbette',
'Bobbi',
'Bobbie',
'Bobby',
'Bobette',
'Bobina',
'Bobine',
'Bobinette',
'Bonita',
'Bonnee',
'Bonni',
'Bonnie',

'Bonny',
'Brana',
'Brandais',
'Brande',
'Brandea',
'Brandi',
'Brandice',
'Brandie',
'Brandise',
'Brandy',
'Brea',
'Breanne',
'Brear',
'Bree',
'Breena',
'Bren',
'Brena',
'Brenda',
'Brenn',
'Brenna',
'Brett',
'Bria',
'Briana',
'Brianna',
'Brianne',
'Bride',
'Bridget',
'Bridgett',
'Bridgette',
'Bridie',
'Brier',
'Brietta',
'Brigid',
'Brigida',
'Brigit',
'Brigitta',
'Brigitte',
'Brina',
'Briney',
'Briny',
'Brit',
'Brita',
'Britaney',
'Britani',
'Briteny',
'Britney',
'Britni',
'Britt',
'Britta',
'Brittan',
'Brittany',
'Britte',
'Brittney',
'Brook',
'Brooke',
'Brooks',
'Brunella',
'Brunhilda',
'Brunhilde',
'Bryana',
'Bryn',
'Bryna',
'Brynn',
'Brynna',
'Brynne',

'Buffy',
'Bunni',
'Bunnie',
'Bunny',
'Burta',
'Cabrina',
'Cacilia',
'Cacilie',
'Caitlin',
'Caitrin',
'Cal',
'Calida',
'Calla',
'Calley',
'Calli',
'Callida',
'Callie',
'Cally',
'Calypso',
'Cam',
'Camala',
'Camel',
'Camella',
'Camellia',
'Cameo',
'Cami',
'Camila',
'Camile',
'Camilla',
'Camille',
'Cammi',
'Cammie',
'Cammy',
'Canada',
'Candace',
'Candi',
'Candice',
'Candida',
'Candide',
'Candie',
'Candis',
'Candra',
'Candy',
'Cappella',
'Caprice',
'Cara',
'Caralie',
'Caren',
'Carena',
'Caresa',
'Caressa',
'Caresse',
'Carey',
'Cari',
'Caria',
'Carie',
'Caril',
'Carilyn',
'Carin',
'Carina',
'Carine',
'Cariotta',
'Carissa',
'Carita',
'Caritta',

'Carla',
'Carlee',
'Carleen',
'Carlen',
'Carlena',
'Carlene',
'Carley',
'Carli',
'Carlie',
'Carlin',
'Carlina',
'Carline',
'Carlisle',
'Carlita',
'Carlota',
'Carlotta',
'Carly',
'Carlye',
'Carlyn',
'Carlynn',
'Carlynnne',
'Carma',
'Carmel',
'Carmela',
'Carmelia',
'Carmelina',
'Carmelita',
'Carmella',
'Carmelle',
'Carmen',
'Carmina',
'Carmine',
'Carmita',
'Carmon',
'Caro',
'Carol',
'Carol-Jean',
'Carola',
'Carolan',
'Carolann',
'Carole',
'Carolee',
'Caroleen',
'Carolie',
'Carolyn',
'Carolina',
'Caroline',
'Caroljean',
'Carolyn',
'Carolyne',
'Carolynn',
'Caron',
'Carree',
'Carri',
'Carrie',
'Carrissa',
'Carrol',
'Carroll',
'Carry',
'Cary',
'Caryl',
'Caryn',
'Casandra',
'Casey',
'Casi',

'Casia',
'Casie',
'Cass',
'Cassandra',
'Cassandre',
'Cassandry',
'Cassaundra',
'Casse',
'Cassi',
'Cassie',
'Cassondra',
'Cassy',
'Cat',
'Catarina',
'Cate',
'Caterina',
'Catha',
'Catharina',
'Catharine',
'Cathe',
'Cathee',
'Catherin',
'Catherina',
'Catherine',
'Cathi',
'Cathie',
'Cathleen',
'Cathlene',
'Cathrin',
'Cathrine',
'Cathryn',
'Cathy',
'Cathyleen',
'Cati',
'Catie',
'Catina',
'Catlaina',
'Catlee',
'Catlin',
'Catrina',
'Catriona',
'Caty',
'Cayla',
'Cecelia',
'Cecil',
'Cecile',
'Ceciley',
'Cecilia',
'Cecilla',
'Cecily',
'Ceil',
'Cele',
'Celene',
'Celesta',
'Celeste',
'Celestia',
'Celestina',
'Celestine',
'Celestyn',
'Celestyna',
'Celia',
'Celie',
'Celina',
'Celinda',
'Celine',

'Celinka',
'Celisse',
'Celle',
'Cesya',
'Chad',
'Chanda',
'Chandal',
'Chandra',
'Channa',
'Chantal',
'Chantalle',
'Charil',
'Charin',
'Charis',
'Charissa',
'Charisse',
'Charita',
'Charity',
'Charla',
'Charlean',
'Charleen',
'Charlena',
'Charlene',
'Charline',
'Charlot',
'Charlott',
'Charlotta',
'Charlotte',
'Charmain',
'Charmaine',
'Charmane',
'Charmian',
'Charmine',
'Charmion',
'Charo',
'Charyl',
'Chastity',
'Chelsae',
'Chelsea',
'Chelsey',
'Chelsie',
'Chelsy',
'Cher',
'Chere',
'Cherey',
'Cheri',
'Cherianne',
'Cherice',
'Cherida',
'Cherie',
'Cherilyn',
'Cherilynn',
'Cherin',
'Cherise',
'Cherish',
'Cherlyn',
'Cherri',
'Cherrita',
'Cherry',
'Chery',
'Cherye',
'Cheryl',
'Cheslie',
'Chiarra',
'Chickie',

'Chicky',
'Chiquita',
'Chloe',
'Chloette',
'Chloris',
'Chris',
'Chriss',
'Chrissa',
'Chrissie',
'Chrissy',
'Christa',
'Christabel',
'Christabella',
'Christabelle',
'Christal',
'Christalle',
'Christan',
'Christean',
'Christel',
'Christen',
'Christi',
'Christian',
'Christianana',
'Christiane',
'Christie',
'Christin',
'Christina',
'Christine',
'Christy',
'Christyna',
'Chrysa',
'Chrysler',
'Chrystal',
'Chryste',
'Chrystel',
'Ciara',
'Cicely',
'Cicily',
'Ciel',
'Cilka',
'Cinda',
'Cindee',
'Cindelyn',
'Cinderella',
'Cindi',
'Cindie',
'Cindra',
'Cindy',
'Cinnamon',
'Cissie',
'Cissy',
'Clair',
'Claire',
'Clara',
'Clarabelle',
'Clare',
...]

In [5]:

```
#find the number of names in the file
len(names.words())
```

Out[5]:

7944

In [6]:

```
#label the names
labelledName = [(name, 'male') for name in names.words('male.txt')] + [(name, 'female') for
name in names.words('female.txt')]
```

In [7]:

```
random.shuffle(labelledName)
featuresets = [(genderFunc(n),gender) for (n,gender) in labelledName]
```

In [8]:

```
#split the training set and the testing set
train,test = featuresets[5000:],featuresets[:2000]
classifier = nltk.NaiveBayesClassifier.train(train)
```

In [9]:

```
classifier.classify(genderFunc('Sakshi'))
```

Out[9]:

'female'

In [10]:

```
classifier.classify(genderFunc('Karthik'))
```

Out[10]:

'male'

In [11]:

```
#the accuracy
nltk.classify.accuracy(classifier,test)
```

Out[11]:

0.753

In [12]:

```
from sklearn.feature_extraction.text import CountVectorizer
vect = CountVectorizer(binary=True)

corpus = ["Sisy is a smart and silly girl."]
```

In [13]:

```
vect.fit(corpus)
```

Out[13]:

```
CountVectorizer(analyzer='word', binary=True, decode_error='strict',
                dtype=<class 'numpy.int64'>, encoding='utf-8', input='content',
                lowercase=True, max_df=1.0, max_features=None, min_df=1,
                ngram_range=(1, 1), preprocessor=None, stop_words=None,
                strip_accents=None, token_pattern='(?u)\\b\\w\\w+\\b',
                tokenizer=None, vocabulary=None)
```

In [14]:

```
vocab = vect.vocabulary_
```

In [15]:

```
for key in sorted(vocab.keys()):
    print("{}:{}".format(key, vocab[key]))
```

```
ancestors:0
are:1
filntstones:2
jetsons:3
of:4
the:5
```

In [16]:

```
vect.transform(["This is a good optical illusion"]).toarray()
```

Out[16]:

```
array([[0, 0, 0, 0, 0, 0]])
```

In [20]:

```
#find the similarity
from sklearn.metrics.pairwise import cosine_similarity
similarity = cosine_similarity(vect.transform(["smart and silly girl"]).toarray(), vect.trans
form(["smart girl who is still silly"]).toarray())
similarity
```

Out[20]:

```
array([[0.]])
```

In []: