

```
#seeing about naivebayes classifiers
```

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

```
gender_features('Samay')
```

```
{'last_letter': 'y'}
```

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

```
names.words()
```

```
['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',
'Agnes',
'Agnese',
'Agnesse',
'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',
'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',
'Bevvv',
'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',
'Carolin',
'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',
'Cassey',

'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',
'Christiana',
'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',  
...]
```

```
print(len(names.words()))
```

```
7944
```

```
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)
```

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

```
train_set, test_set = featuresets[:5000], featuresets[2000:]
```

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

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

```
'female'
```

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

```
0.7639636608344549
```

```
#####  
#####  
#####
```

```
#####  
#####  
#####
```

```
import nltk
```

```
f = open('tweets1.txt','r')
```

```
text2.concordance("the")
```

Displaying 25 of 349 matches:

o bed. gary coleman sighting by Zoom. The rumors were true Festmob. In
The Sign

oom. The rumors were true Festmob. In The Signal premiere. Lotsa
buyers. Good l

a buyers. Good luck gang listening to the podcast net@nite while
setting up twit

looking for family Holy holy holy is the lord God almighty Wow she
finally fin

ing in longgggggggggg line for food, the only question is what to
eat? Lunch t

!!!!!! Lol technical difficulties and the crowd is singing for joy

Main session

o find people I am sitting waitng for the session.....I HATE NEGATIVE
SPACE...

i released in Singapore yet?! surfing the net playing around on
twitter At work

ng around on twitter At work Watching the valentine's day pillowfight
sans pill

ere and not here. Talking about 24 in the cars, we pass Bauer's limo
service. A

ack Morass. Getting better! Woo, felt the tremor. I feeling giddy!
heading to b

n for let url Listening to music with the kids url looking at social
network si

o p2p to sourceforge url listening to the Gossip and Arctic Monkeys on
Hype Mac

spreadsheet for Building Surveyors at the British Red cross Society
probably ta

. url Going to a meeting on Israel at the British Red Cross Adding a

filter to
y on Rails spreadsheet. Going home on the overground Babbling,
seemingly homele
y homeless San Francisco bag lady all the more disconcerting for
carrying Sur L
take along some of those glasses with the eyes painted on... Sitting
on the cou
ith the eyes painted on... Sitting on the couch, reading email
Seriously jonesi
in' for Nairns mini-oatcakes---verily the heroin of Scotland!
coffeebreak :-) A
otland! coffeebreak :-) Astonished by the rain gusting over the city .
.where d
) Astonished by the rain gusting over the city . .where did Saturday's
lemonade
shine go? Of course my perspective on the weekend weather was warped
by hiking
ails spreadsheet that I'm writing for the BRCS (url). 6pm: Living
Black invest
a big outback party in Hermannsburg. THE MIGHTY BOOSH, 10:05pm:
Howard hooks a

```
from urllib import request
url = "https://en.wikipedia.org/wiki/Tiger"
response = request.urlopen(url)
raw = response.read().decode('utf8')
type(raw)
str
len(raw)
845063
raw[:75]
'<!DOCTYPE html>\n<html class="client-nojs vector-feature-language-in-
header- '
from nltk.tokenize import word_tokenize
tokens = word_tokenize(raw)
type(tokens)
list
```

HTML => ASCII => TEXT => VOCAB

Cell In[85], line 1

```
HTML => ASCII => TEXT => VOCAB
      ^
```

SyntaxError: invalid syntax