

```
In [1]: #import Libraries
import sqlite3
import nltk
import random
import numpy as np
from nltk.corpus import stopwords
from nltk.probability import FreqDist
from string import punctuation
from collections import Counter, defaultdict

# Some punctuation variations
punctuation = set(punctuation) # speeds up comparison
tw_punct = punctuation - {"#"}

# Stopwords
#from textbook
sw = set(nltk.corpus.stopwords.words('english'))

def remove_stop(text) :
    # modify this function to remove stopwords
    #from textbook
    #return [t for t in tokens if t.lower() not in sw]
    words = text.split()
    result_str = ""
    for word in words:
        if(word not in sw):
            result_str = result_str + " " + word
    return(result_str)

def remove_punctuation(text, punct_set=tw_punct) :
    return("".join([ch for ch in text if ch not in punct_set]))

def tokenize(text) :
    """ Splitting on whitespace rather than the book's tokenize function. That
        function will drop tokens like '#hashtag' or '2A', which we need for Twitter. """

    # modify this function to return tokens
    text = text.split()
    return(text)

def prepare(text, pipeline) :
    tokens = str(text)

    for transform in pipeline :
        tokens = transform(tokens)

    return(tokens)
```

```
In [2]: convention_db = sqlite3.connect("C:/Users/Abana/Downloads/2020_Conventions.db")
print(convention_db)
convention_cur = convention_db.cursor()
print(convention_cur)
```

```
<sqlite3.Connection object at 0x0000010C37421E30>
```

```
<sqlite3.Cursor object at 0x0000010C60A1E570>
```

## Part 1: Exploratory Naive Bayes

We'll first build a NB model on the convention data itself, as a way to understand what words distinguish between the two parties. This is analogous to what we did in the "Comparing Groups" class work. First, pull in the text for each party and prepare it for use in Naive Bayes.

In [4]:

```

convention_data = []
# fill this list up with items that are themselves lists. The
# first element in the sublist should be the cleaned and tokenized
# text in a single string. The second element should be the party.

query_results = convention_cur.execute(
    "SELECT text, party FROM conventions"
)
my_pipeline = [str.lower]#, remove_punctuation, tokenize, remove_stop]

for row in query_results:
    # store the results in convention_data
    text, party = row
    text = str.lower(text)
    text = remove_punctuation(text)
    #text = tokenize(text)
    text = remove_stop(text)
    #apply(prepare, pipeline=my_pipeline)
    convention_data.append((text, party))

...

s = "skip and the to content the company careers press freelancers blog x services tran
words = s.split()
result_str = ""
for word in words:
    if(word not in sw):
        result_str = result_str + " " + word
print(result_str)
print()
...

#my_pipeline = [str.lower, remove_punctuation, tokenize, remove_stop]
#convention_data["tokens"] = convention_data[text].apply(prepare, pipeline=my_pipeline)

print(convention_data[0])

```

```

(' skip content company careers press freelancers blog x services transcription captions
foreign subtitles translation freelancers contact login « return transcript library home
transcript categories transcripts 2020 election transcripts classic speech transcripts c
ongressional testimony hearing transcripts debate transcripts donald trump transcripts e
ntertainment transcripts financial transcripts interview transcripts political transcrip
ts press conference transcripts speech transcripts sports transcripts technology transcr
ipts aug 21 2020 2020 democratic national convention dnc night 4 transcript rev > blog >
transcripts > 2020 election transcripts > 2020 democratic national convention dnc night
4 transcript night 4 2020 democratic national convention dnc august 20 read full transcr
ipt event transcribe content try rev free save time transcribing captioning subtitling',
'Democratic')

```

In [5]:

```
random.choices(convention_data,k=10)
```

```

Out[5]: [(' would please rise playing national anthem', 'Republican'),
(' know took good long time get hope trump wins continue progress he's made policies h
e's put place let's get economy back track let's moving right direction',
'Republican'),

```

```

(' they're right next room great night julia', 'Democratic'),
(' discrimination denied many black americans chance home build wealth joe Biden plan e
nd racist lending practices help people color achieve american dream owning home isn't r
acial justice it's strong communities economic security working families illinois casts
59 votes Bernie Sanders 122 votes next president Joe Biden',
'Democratic'),
(' today channel warrior roots battling human trafficking i've able lead largest traffi
cking prosecutions America traveled foreign countries working law enforcement NGOs disma
ntle trafficking networks rescue people brutal conditions imaginable young girls women s
old sex slavery young boys men forced labor servitude illegal adoptions organ harvesting
human life utterly debased',
'Republican'),
(' finally rock getting hours weeks months unspeakable pain unending surgeries anchor r
elearned walk helping every step every stumble military spouses hold families together p
raying loved ones safety wherever they're deployed serving caregivers disabled service m
embers picking pieces starting whenever next tour next war arises Joe Biden understands
sacrifices he's made son Beau deployed Iraq burden also shouldered family Joe knows fear
military families live he's felt dread never knowing deployed loved one safe understands
bravery muster strength every hour every day Beau overseas that's kind leader service me
mbers deserve',
'Democratic'),
(' good evening i'm senator Tim Scott great state South Carolina tuning participating p
olitical process God bless isn't pictured tonight country experiencing something none us
envisioned global pandemic deaths George Floyd Breonna Taylor 2020 tested nation ways ha
ven't seen decades regardless challenges presented us every four years Americans come to
gether vote share stories makes nation strong lessons learned strengthen generations ele
ction Donald Trump Joe Biden solely Donald Trump Joe Biden it's promise America it's cha
llenges heartbreaks hopes dreams it's respond tackling critical issues like police refor
m sen',
'Republican'),
(' somebody believes what's best us someone like Joe Biden actually believes American i
dea that's kind person want White House',
'Democratic'),
(' morning last election said owe Donald Trump open mind chance lead meant every presid
ent deserves Trump came much set strong economy plans managing crises including pandemic
yes Democrats would disagreed many things put interests ego aside seeing humanity child
ripped parents border protestor calling justice family wiped natural disaster would good
thing America world wish Donald Trump knew president America needs president right throu
ghout time crisis Americans keep going checking neighbors showing jobs first responders
hospitals grocery stores nursing homes yes still takes village need leaders equal moment
sacrifice service',
'Democratic'),
(' good evening i'm governor Kristi Noem South Dakota i'm tonight believe America excep
tional nation founded three principles equality freedom opportunity today founding princ
iples attack year choice Americans man values ideals built man isn't guided ideals coinc
identally built nothing remember America's battle independence fight selfgovernance some
thing never done men great intellect wisdom like James Madison father constitution hoped
constitutional republic would last ages mitigate problems would naturally arise politica
l factions prevent tyranny Madison also authored much bill rights understood natural ten
dency government increasingly encroach people's consent thus freedom',
'Republican'))]
```

In [6]:

```

word_cutoff = 5

tokens = [w for t, p in convention_data for w in t.split()]

word_dist = nltk.FreqDist(tokens)
feature_words = set()

for word, count in word_dist.items():
    if count > word_cutoff:
        feature_words.add(word)
```

```
print(feature_words)
#print(f"With a word cutoff of {word_cutoff}, we have {len(feature_words)} as features
```

```
{'tv', 'cast', 'church', 'together', 'roads', 'happened', 'never', 'goes', 'responders',
'invest', 'christ', 'unprecedented', 'informed', 'problem', 'words', 'risk', 'wait', 'se
rve', 'israel', 'factory', 'mattered', 'happening', 'schumer', 'party's', 'please', 'b
s', 'walk', 'activist', 'mchenry', 'address', 'nursing', 'portland', 'heads', 'matter',
'grant', 'system', 'chosen', 'denied', 'wife', 'arrested', 'doesn't', 'pay', 'violence',
'100', 'left's', 'march', 'captioning', 'giving', 'starting', 'presidential', 'less', 'a
pp', 'oregon', 'forever', 'countless', 'abraham', 'minnesota', 'consider', 'talking', 't
ell', 'team', 'jack', 'ask', 'changes', "if", 'soldiers', 'poor', 'interview', 'wants',
'backgrounds', 'code', 'heroic', 'choosing', 'talked', 'range', 'chair', 'telling', 'ene
mies', 'want', 'sat', 'service', 'brings', 'champion', 'self', 'alexandria', 'politician
s', 'didn't', 'equality', 'protecting', 'centers', 'indivisible', 'endure', 'pastor', 'c
onstruction', 'wake', 'reason', 'real', 'got', 'floor', 'upon', 'ignored', 'social', 'in
troduce', 'violent', 'role', 'see', 'uniform', 'climate', 'thrive', 'must', 'behind', 's
ister', 'innocent', 'talk', 'motto', 'years', 'america', 'accountable', 'canceled', 'sci
ence', 'listened', 'room', 'despite', 'worship', 'registered', 'believed', 'experiencin
g', 'absolutely', 'chance', 'proven', 'still', 'harris', 'market', 'tear', 'quote', 'har
der', 'soleimani', 'endless', 'victims', 'state', 'madison', 'believe', 'stage', 'proble
ms', 'civil', 'determine', 'clearly', 'email', 'involved', 'vote', 'doctors', 'option',
'threatened', 'order', 'pain', 'earlier', 'joe', 'entertainment', '10', '3rd', 'second',
'worst', 'losing', 'speeches', 'lincoln's', 'restaurants', 'change', 'terrorists', 'trad
e', 'heroes', 'hospitals', 'along', 'expected', 'celebrate', 'table', 'cuts', 'protest',
'current', 'felt', 'crosstalk', 'struggling', 'held', 'include', 'word', 'translation',
'belief', 'many', 'everything', 'threat', 'smile', 'wanted', 'something', 'plans', 'prom
ote', 'troops', 'immigrant', 'nature', 'blue', 'saved', 'decision', 'hospital', 'human',
'calls', 'approved', 'thousands', 'basic', 'masks', 'bill', 'career', 'income', 'follo
w', 'equally', 'francisco', 'postal', 'peacefully', 'ii', 'scranton', 'three', 'critica
l', 'moonshot', 'communism', 'thank', 'albaghdadi', 'michigan', 'abroad', 'highest', 'h
alf', 'year's', 'mine', 'cross', 'boundless', 'families', 'classroom', 'foreign', 'moo
n', 'hoax', 'mail', 'gratitude', 'true', 'fund', 'task', 'they've', 'others', 'though
t', 'advice', 'frontline', 'truth', 'hi', 'total', 'authority', 'can't', 'move', 'gradu
ated', 'ban', 'congressman', 'relief', 'names', 'healing', 'wisconsin', 'private', 'fun
ding', 'young', 'officials', 'whether', 'grieving', 'production', 'led', 'holding', '20
17', 'choose', 'farmers', 'mass', 'watching', 'jobs', 'god's', 'similar', 'weren't', 's
hop', 'billion', 'increased', 'iran', '12', "what", 'waves', 'torch', 'anyone', 'raisi
ng', 'promises', 'need', 'safely', 'september', 'illegal', 'might', 'children's', 'gir
l', 'support', 'whole', 'north', 'pelosi', 'c', 'sexual', 'weak', 'amen', 'plan', 'folk
s', 'faces', '27', 'showed', 'environmental', 'terrible', 'hat', 'say', 'open', '70',
'office', 'biden', 'looked', 'delaware', 'don't', 'forces', 'agreed', 'voices', 'auto',
'convention', 'reelection', 'football', 'full', 'john', 'usa', 'quiet', 'representative
s', 'peaceful', 'mom', '2008', 'women's', 'classic', 'allies', 'examples', 'recently',
'liberty', 'hearts', 'shift', 'secure', 'markets', 'union', 'impacted', 'certain', 'war
s', 'sick', 'loves', 'invisible', 'use', 'donald', 'deny', 'health', 'aoc', 'spent', 's
ought', 'crush', 'voter', 'lincoln', 'racial', 'bernie', 'historically', 'capable', 'ma
rginalized', 'karen', 'leaders', 'place', 'legislation', 'reelect', 'remember', 'stee
l', 'special', 'truly', 'things', 'radical', '2009', 'citizenship', 'group', 'said', 'w
elcome', 'value', 'congresswoman', 'meaningful', 'votes', 'putting', 'taken', 'live',
'dream', 'hardworking', 'placed', 'traveled', 'turkey', 'islands', 'dangerous', 'gentle
men', 'tremendous', 'border', 'subtitles', 'doubt', 'april', 'supporters', 'achieve',
'salute', '2019', 'empty', 'lifesaving', 'prevent', 'country', 'decent', 'massive', 'se
en', 'improve', 'removed', 'thriving', 'anymore', 'segregation', 'advantage', 'respect
s', 'board', 'poverty', 'heart', 'radio', 'skip', 'wealth', 'supposed', 'men', 'progres
sive', 'wage', 'dave', 'continuing', 'corps', 'stop', 'granted', 'spoke', 'workers', 'e
ducated', 'best', 'suffer', 'baby', 'strong', 'country's', 'achieved', 'w', 'leading',
'reflected', 'could', 'sure', 'governor', 'walking', 'toward', 'without', 'landmark',
'closer', 'says', 'gets', 'americans', 'pretty', 'sitting', 'come', 'silenced', 'traini
ng', 'financial', 'restore', 'schools', 'continues', 'technology', 'broke', 'big', 'yes
terday', 'hopes', 'neighbors', 'subtitling', 'sir', 'call', 'hostage', 'allowed', 'burd
en', 'grandmother', 'communities', 'robert', '5603', 'obama's', 'optimism', 'fear', 'fa
vorite', '48', 'cause', 'tirelessly', 'followed', 'souls', 'content', 'sunday', 'mothe
r's', 'san', 'mayor', 'feels', 'values', 'tells', 'eye', 'interested', 'left', 'comes',
'trying', 'afford', 'sacrifice', 'texas', 'irish', 'nominating', 'cases', 'republican
s', 'earth', 'x', 'you're', 'mostly', 'decade', 'nice', 'debt', '80', 'cover', 'prisone
```

rs', 'singing', 'standing', 'morning', 'greater', 'passing', 'supply', 'adversaries', 'continue', 'recovered', 'facts', 'seemed', 'walked', 'six', 'chaos', 'marine', 'army', 'moving', 'afraid', 'primary', '65', 'river', 'neighborhoods', 'point', 'launched', 'arms', 'waiting', 'police', 'transcribe', 'killing', 'phenomenal', 'isn't', 'privilege', 'moral', 'builds', 'oh', 'nine', 'respected', 'asian', 'unified', 'regardless', 'bed', 'name', 'indian', 'medical', 'montana', 'heaven', 'lands', 'constant', 'uphold', 'played', 'fairer', 'p', 'trouble', 'ground', 'onto', 'square', 'back', 'stayed', 'fighting', 'hurt', 'protestors', 'europe', '40', 'history', 'free', 'wall', 'darkness', 'mean', 'obambabiden', 'century', 'bear', 'liberties', 'speak', 'resources', 'daughter', 'patients', 'falls', 'spread', 'afghanistan', 'places', 'member', 'forgotten', 'judges', 'correct', 'concerned', 'election', 'question', 'rid', 'council', 'ancestors', 'bible', 'heard', 'fast', 'small', 'condition', 'justice', 'recovery', 'borders', '•', 'exactly', 'wins', 'pushing', 'south', '60', 'regular', 'signed', 'exceptional', 'heavy', 'supporting', 'throughout', 'compassionate', 'thinking', 'allowing', 'represent', 'side', 'defense', 'center', 'gave', 'won't', 'declaration', 'decide', 'realized', 'executive', 'anything', 'murder', 'perform', 'religious', 'proud', 'would', 'sacrificed', 'it', 'experience', 'taught', 'businesses', 'rather', 'capacity', 'tennessee', 'browser', 'jump', 'devotion', 'hoped', 'obligation', 'grocery', 'brown', 'undo', 'api', 'democrat', 'parties', 'veteran', 'knew', 'help', 'agenda', 'low', 'lives', 'majority', 'minority', 'security', 'here's', 'tape', 'jr', 'whose', '8883690701', 'stay', 'housing', 'we're', 'immediately', 'pursue', 'behalf', 'largest', 'mistake', 'pulled', 'moment', 'world', 'particular', 'remind', 'ego', 'doors', 'freelancers', 'judge', 'o'rourke', 'systemic', 'mitch', 'tragedy', 'defunding', 'sides', 'cost', 'guy', 'eyes', 'overcome', 'choice', 'awarded', 'designed', 'define', 'even', 'pray', 'defeat', 'king', 'went', 'makes', 'testament', 'iraq', 'invite', 'situation', 'duty', 'coming', 'agree', 'rising', 'chinese', 'confidence', 'witness', 'turning', 'hatred', 'wonderful', 'submit', 'native', 'andrew', 'areas', 'soon', 'diversity', 'city', 'growing', 'new', 'sons', 'prices', 'beloved', 'always', 'independents', 'transcripts', 'amazing', 'brought', 'ahead', 'asked', 'online', 'meet', 'blessed', 'become', 'pass', '19th', 'done', 'transcribing', 'roll', 'pride', 'kearny', 'day', 'treated', 'solve', 'successful', 'rescue', 'yeah', 'jerusalem', 'sit', 'given', 'year', 'virtually', 'anniversary', 'exciting', 'race', 'going', 'think', 'there's', 'forget', 'reach', 'mind', 'entire', 'wearing', 'oath', 'regime', 'certainly', 'around', 'prosperous', 'news', 'weakness', 'including', 'forced', 'lie', 'armed', 'sadly', 'sins', 'education', 'homes', 'air', 'grow', 'captions', 'bridges', 'save', 'released', 'fit', 'greetings', 'effective', 'shoulders', 'shown', 'grandparents', 'prayers', 'middle', 'demands', 'personal', 'diverse', 'limited', 'lewis', 'jordan', 'pledge', 'prayer', 'chicago', 'deliver', 'began', 'responded', 'trafficking', 'celebrating', 'hampshire', 'socialism', 'sisters', 'j', 'promise', 'testing', 'grace', 'within', 'cory', 'decades', 'african', 'made', 'wear', 'effects', '@', 'memory', 'information', 'organizations', 'lawlessness', 'mothers', 'fail', '-', 'guard', 'recognized', 'devastating', 'cold', 'winning', '25', 'stands', 'bless', 'keep', 'hardship', 'ronald', 'play', 'toughest', 'rioting', 'hearing', 'ideals', 'i', 'today', 'couldn't', 'clinton', 'destroyed', 'another', 'enjoy', 'sen', 'appointed', 'marched', 'strategy', 'cops', 'friendship', 'preexisting', 'begin', 'that's', 'train', 'humility', 'assault', 'profound', 'overseas', 'mexico', 'prescription', 'inclusive', 'high', 'touch', 'working', 'came', 'smart', 'august', 'read', 'towards', 'married', 'dad', '18', 'kentucky', 'raise', 'evil', 'it", 'november', 'dreams', 'firsthand', 'philadelphia', '>', 'tom', 'bidenharris', 'rule', 'good', 'impossible', 'she's', 'b', 'following', 'shouldn't', 'converted', 'long', 'unborn', 'indiana', 'wow', 'struck', 'advocate', '1', 'coronavirus', 'empathy', 'fall', 'i've', 'i'm', 'friends', 'utah', 'several', 'filled', 'congratulations', 'track', 'recent', 'secretary', 'mcconnell', 'floyd', 'sight', 'last', 'saving', 'overwhelming', 'else's', 'sound', 'illness', 'sally', 'reclaim', 'chapter', 'beacon', 'immigrants', 'criminals', '<', 'trust', 'bigger', 'helps', 'laid', 'parent', 'deep', 'covid19', 'opponents', 'possibility', 'nominee', 'count', 'two', 'top', 'leadership', 'enemy', 'beliefs', 'seniors', 'girls', 'knows', 'latina', 'watched', 'rev', 'inspiration', 'rest', 'written', 'inaudible', 'journey', 'text', 'committed', 'case', 'embrace', 'nevada', 'louisiana', 'mighty', 'ca', 'cancel', 'you'll', 'understands', 'belong', 'community', 'people', 'pushed', 'riots', 'law', 'prison', 'paycheck', 'grew', 'mrs', 'residents', 'responsible', 'volunteer', 'greatness', 'field', 'basically', 'box', 'sacred', 'senior', 'bridge', 'kindness', '17', 'unity', 'terrorist', 'raised', 'issue', 'advanced', 'attacked', 'lady', 'hands', 'divides', 'known', 'priority', 'presidency', 'melania', '19', 'programs', 'running', '16', 'spouses', '3', 'stepped', 'relentless', 'struggles', 'available', 'greatest', 'broken', 'so-called', 'party', 'effort', 'numbers', 'purpose', 'blame', 'blog', 'served', 'hundreds', 'head', 'get', 'streets', 'infrastructure', 'happy', 'progress', 'nancy', 'invited', 'defined', 'tackle', 'ones', 'child', 'uni

que', 'ripped', 'east', 'california', 'praise', 'paid', 'return', 'caused', 'halls', 'b  
 lames', 'nation', 'government', 'roosevelt', 'cash', 'who's', 'idea', 'hopeful', 'unlea  
 shed', 'crime', 'prosperity', 'figure', 'nation's', 'clear', 'united', 'joe Bidencom',  
 'acting', 'possibly', 'called', 'uncertain', 'talented', 'ideas', 'federal', '94108',  
 'dedicated', 'reform', 'family's', 'report', 'fix', 'administration', 'unite', 'electio  
 ns', 'everyday', 'speed', 'modern', 'scott', 'hell', 'ja'ron', 'love', 'confront', 'wou  
 ldn't', 'worry', 'learned', 'inspires', 'hold', 'immigration', 'named', '38', 'option  
 s', 'different', 'conviction', '13', 'faith', '22', 'chief', 'bold', 'look', 'determine  
 d', 'miracle', 'future', 'forward', 'became', 'secured', 'showing', 'enable', 'early',  
 'decisions', 'socialist', 'prepared', 'casts', 'student', 'ticket', 'evening', 'built',  
 'drug', 'cut', 'affected', 'based', 'tears', 'feel', 'failed', '24', 'supreme', 'high  
 r', 'leave', 'reminded', 'worse', 'better', 'action', 'potential', 'enforcement', 'di  
 e', 'anywhere', 'lot', 'iowa', 'generations', 'chuck', 'destroying', 'success', 'mean  
 t', 'he'd', 'development', 'rebuilt', 'hurricane', 'july', 'everyone', 'everyone's', 'p  
 rogram', 'republic', 'cherish', 'movement', '2016', 'earned', 'confronted', 'died', 'to  
 night', 'know', 'vaccine', 'tax', 'black', 'goal', 'views', 'test', 'ultimately', 'sou  
 l', 'well', 'ready', 'elizabeth', 'scared', 'boy', 'fights', 'corporations', 'met', 'wa  
 rrior', 'results', 'freely', 'awesome', 'grandfather', 'zones', 'division', 'ventilator  
 s', 'offer', 'happens', 'video', 'form', 'resilient', 'sports', 'honest', 'media', 'ens  
 ure', 'listens', 'mike', 'ran', 'due', 'hate', 'safer', 'departments', 'login', 'lear  
 n', 'yes', 'individual', 'nato', 'activists', 'women', 'protests', 'thinks', 'points',  
 'conservative', 'washington', 'tests', 'also', 'apart', 'suffrage', 'local', 'labor',  
 'complete', '47', 'view', 'criminal', 'difference', 'direction', 'teach', 'protective',  
 'bringing', 'voters', 'digest', 'voting', 'nations', 'ignore', 'quit', 'almost', 'fairn  
 ess', 'jersey', 'work', 'site', 'millions', 'weeks', 'looks', 'loss', 'took', 'weekly',  
 'employees', 'produce', 'silence', 'whatever', 'practice', 'senator', 'loving', 'apprec  
 iate', 'political', 'grown', 'watch', 'way', 'taking', 'failure', 'give', 'transcript',  
 'wrong', 'turn', 'fire', 'process', 'nothing', 'almighty', 'incredible', 'we've', 'char  
 ge', 'they're', 'major', 'art', 'lucky', 'founding', 'healthcare', 'emma', 'week's', 'm  
 oms', 'provide', 'town', 'shocked', 'courageous', 'ways', 'growth', 'classrooms', 'piec  
 es', 'sending', 'stood', 'light', 'responsibility', 'recover', 'stronger', 'none', 'pro  
 vided', 'story', 'food', 'right', 'patriots', 'who've', 'one', 'required', 'i'll', 'off  
 icers', 'leads', 'communist', 'kamala', 'run', 'answer', 'reflect', 'hour', 'tough', 'n  
 eeded', 'trump', 'sanders', 'shutdown', 'else', 'jon', 'conversation', 'hundred', 'ris  
 e', 'wing', 'actually', 'restored', 'laura', 'enacted', 'storm', 'inspiring', 'maria',  
 'founders', 'injustice', 'reasons', 'meaning', 'mask', 'issues', 'preserve', 'hardest',  
 'ability', 'emergency', 'bought', '21st', 'american', 'establishment', 'maine', 'star',  
 'lee', 'considered', 'generation', 'sees', 'listening', 'praying', 'farmer', 'countrie  
 s', 'advance', 'cares', 'ivanka', 'level', 'check', 'you"', 'revcom', 'victory', 'colle  
 ges', 'careers', 'grade', 'florida', 'rebuild', 'longer', 'freedoms', 'heights', 'awa  
 y', 'passion', 'aren't', 'department', 'put', 'latino', 'maybe', 'cannot', 'deliverin  
 g', 'suffering', 'witnessed', 'okay', 'survived', 'fortunate', 'life', 'least', 'challe  
 nged', 'stories', 'tools', 'services', 'voice', 'officer', 'final', 'telehealth', 'dark  
 est', 'used', 'clean', 'job', 'skin', 'district', '2020', 'defend', 'shaped', 'cochai  
 r', 'signing', 'lifted', 'skills', 'we'd', 'driven', 'banks', 'director', 'updated', 'a  
 ge', 'manage', 'ago', 'conference', 'deployed', 'returned', 'parenthood', 'lessons', 'c  
 ommittee', 'independent', 'charged', 'holds', 'battle', 'kids', 'probably', 'forth', 's  
 imple', 'importantly', 'youngest', 'hit', 'million', 'environment', 'home', 'act', 'cat  
 egories', 'hole', 'late', 'thoughts', 'tariffs', 'religion', 'killed', 'corner', 'bod  
 y', 'created', 'dear', 'attorney', 'school', 'rights', 'prove', 'radicals', 'expand',  
 'spirit', 'receive', 'serves', '4', 'democrats', 'virus', 'months', 'works', 'bright',  
 'korea', 'humanity', 'handle', 'oklahoma', 'supported', 'coverage', 'serving', 'night',  
 'positive', 'literally', 'since', 'alone', 'worried', '2012', 'integrity', 'rural', 'da  
 ily', 'saw', 'yet', '...', 'silent', 'heroism', 'joseph', 'death', 'finally', 'shut', 'la  
 ws', 'louis', 'positions', 'tyranny', 'tired', 'virginia', 'solutions', 'seem', 'speake  
 r', 'treatment', 'minneapolis', 'near', 'planned', 'dnc', 'inbox', 'fellow', 'states',  
 'revolution', '30330', 'attacks', 'jill', 'gather', 'saying', 'drugs', 'protection', 'o  
 wners', 'brothers', 'courage', 'father', 'dark', 'decided', 'post', 'stopped', 'bring',  
 'they'll', 'deadly', 'necessary', 'moments', 'record', 'tomorrow', 'chip', 'medicine',  
 'old', 'investing', 'great', 'chris', 'front', 'wisdom', 'deals', 'puts', 'female',  
 '5', 'send', 'path', 'steady', 'calling', 'request', 'share', 'unfair', 'terms', 'spen  
 d', 'stuck', 'blind', 'economy', 'caption', 'perfect', 'average', 'indigenous', 'perso  
 n', 'close', 'vital', 'space', 'mississippi', 'society', 'races', 'step', 'manufacturin  
 g', 'four', 'what's', 'significant', 'we'll', 'fair', 'hasn't', 'language', 'policy',  
 'wasn't', 'quickly', 'efforts', 'normal', 'developed', 'voted', 'danger', 'heal', 'oper

ation', 'aisle', 'syria', 'kept', 'lies', 'flames', 'third', 'fired', 'faced', 'reflect  
 s', 'travel', 'narrative', 'ambassador', 'challenge', 'abuse', 'husband', 'detective',  
 'failing', 'present', 'beto', 'hillary', 'freedom', 'win', 'murdered', 'beat', 'importa  
 nce', 'dollars', 'hunter', 'northern', 'destiny', 'supporter', 'legacy', 'strongly', 'f  
 riend', 'college', 'keeps', 'force', 'gone', 'lose', 'grandchildren', '911', 'degree',  
 'cure', 'supplies', 'happiness', 'deserve', 'worth', 'mother', 'patriotic', 'island',  
 'principles', 'understanding', 'goodness', 'hours', 'walks', 'term', 'resolve', 'tide',  
 'hill', 'they'd', 'resilience', 'embassy', 'build', 'looting', 'javascript', 'treat',  
 '50', 'democratr', 'door', 'weapons', 'decency', 'strength', 'type', 'effect', 'deser  
 ves', 'it's', 'china's', 'georgia', 'changed', 'company', 'he'll', 'nightmare', 'may',  
 'childcare', 'constitution', '11', 'trump's', 'preparing', 'wish', 'helping', 'unfortun  
 ately', 'wondering', 'aid', 'either', 'establish', 'daughters', 'ill', 'game', 'easie  
 r', 'often', 'leader', 'bad', 'citizens', 'charges', 'rules', 'succeed', 'godgiven', 'l  
 et's', 'commander', 'i'd', 'china', 'domestic', '8th', 'common', 'pursuit', 'democrac  
 y', 'reagan', 'factories', 'stand', 'summer', 'disaster', 'lifetime', 'found', 'educato  
 rs', 'car', 'looking', 'fact', 'organization', 'susan', 'military', 'color', 'república  
 n', 'water', 'threaten', 'extra', 'character', 'institutions', 'national', 'flag', '22  
 2', 'believes', 'remain', 'interest', 'passed', 'ladies', 'cancer', 'audience', 'give  
 s', 'hbcus', 'smith', 'huge', 'anthony', "i'm", 'takes', 'street', 'control', 'underst  
 and', 'brighter', 'iron', 'family', 'speakers', 'colorado', 'international', 'privacy',  
 'press', 'moved', 'healthy', 'believing', '15', 'destroy', 'attack', 'joy', 'show', 'rec  
 ognize', 'rates', 'guess', 'instead', 'attacking', 'lower', 'woman', 'planet', 'former',  
 'beau', 'leaves', 'red', 'reviews', 'rate', 'democratic', 'campaign', '2', 'stuff', 'car  
 ed', 'course', '26', 'beaten', 'cities', 'nobody', 'candidates', 'you'd', 'equipment',  
 'brother', 'slavery', 'across', 'missouri', 'house', 'accident', 'preserving', 'shared',  
 'seat', 'finish', 'message', 'compassion', 'first', 'seems', 'store', 'guests', 'start',  
 'addiction', 'pennsylvania', 'nurses', 'among', 'capital', 'ally', 'tribal', 'powerful',  
 'crisis', 'lowest', 'students', 'willing', 'exercise', 'president's', 'days', 'closely',  
 'imagined', 'sometimes', 'sentence', 'prolife', 'keeping', 'dismantle', 'dead', 'globa  
 l', 'mountain', 'blessing', 'elites', 'defending', 'richard', 'delegates', 'quality', 'g  
 od', 'nuclear', 'divide', 'little', 'martin', 'personally', 'veterans', 'events', 'elect  
 ed', 'members', 'economic', 'tried', 'contact', 'miles', 'sent', 'hear', 'ballot', 'tri  
 p', 'shootings', 'vice', 'hello', 'allow', 'jesus', 'really', 'result', 'grateful', 'eas  
 y', 'safety', 'alongside', 'unlike', 'declare', 'chooses', 'lord', 'billions', 'sad', 'g  
 ift', 'divisions', 'loved', 'vulnerable', 'he's', 'products', 'differences', 'status',  
 'recession', 'biggest', 'turned', 'mob', 'judgment', 'survive', 'speech', 'strongest',  
 'regulations', 'diagnosed', 'bit', 'supports', 'accepted', 'midst', 'response', 'five',  
 'officially', 'someone', 'neighborhood', 'born', 'hero', 'gun', 'supportrevcom', 'hope',  
 'depend', 'interests', 'testimony', 'depression', 'identity', 'benefit', 'allegiance',  
 'fbi', 'example', 'worker', 'children', 'increase', 'experienced', 'individuals', 'congr  
 ess', 'thing', 'mr', 'victim', 'defeated', 'spending', 'everybody', 'using', 'rock', 'co  
 re', 'outside', 'large', 'retired', 'reading', 'find', 'guys', 'shot', 'dignity', 'facin  
 g', 'f', 'ohio', 'remains', 'beautiful', 'promised', 'bags', 'matters', 'every', 'us',  
 'much', 'feed', 'library', 'seven', 'opened', 'unions', 'protected', 'official', 'hard',  
 'farm', 'received', 'wages', 'gender', 'learning', 'mission', 'hey', 'lawyer', 'wheneve  
 r', 'st', 'partner', 'told', 'mile', 'lines', 'western', 'hurting', 'senate', 'county',  
 'guns', 'fort', 'actions', 'respect', 'odds', 'delivered', 'carolina', 'speaking', 'anar  
 chists', 'president', 'consequences', 'conditions', 'samuel', 'sort', 'obama', 'preciou  
 s', 'release', 'enough', 'wonder', 'hostages', 'treating', 'means', 'demonstrated', 'sur  
 vivors', 'counting', 'general', 'presidents', 'kayla', 'ending', 'time', 'crises', 'trad  
 ition', 'historic', 'end', 'focus', 'wrote', 'represents', 'sense', 'able', 'injustice  
 s', 'sounds', 'reopen', 'unless', 'line', 'opportunity', 'extraordinary', 'debate', 'v  
 a', 'month', 'allows', 'shining', 'cherished', 'reduce', 'iowans', 'joe's', 'medicare',  
 'heritage', 'gov', 'abortion', 'companies', 'rioters', 'stake', 'try', 'rose', 'simply',  
 'make', 'everywhere', 'creation', 'george', 'colleagues', 'transcription', 'physical',  
 'world's', 'blood', 'deal', 'systems', 'dakota', 'kill', 'renew', 'candidate', 'overvie  
 w', 'especially', 'happen', 'boys', "the", 'houses', 'beyond', 'teacher', 'court', 'bid  
 en's', 'care', 'policies', 'guarantee', 'natural', 'amendment', 'part', 'congressional',  
 'challenging', 'credentials', 'bills', 'era', 'single', 'honor', 'eight', 'completely',  
 'treatments', 'son', 'coffee', 'barack', 'gas', 'credit', 'certified', 'building', 'raci  
 sm', 'making', 'citizen', 'dr', 'take', 'worthy', 'dairy', 'strengthen', 'fundamental',  
 'kind', 'joined', 'impact', 'inequality', 'equal', 'lift', 'intelligence', 'pandemic',  
 'needs', 'green', 'lost', 'business', 'agreement', 'face', 'grief', 'liberal', 'dc', 'gl  
 ad', 'break', 'pleased', 'challenges', 'talks', 'included', 'russia', 'struggle', 'kid',  
 'expect', 'agriculture', 'class', 'eliminate', 'universities', 'starts', 'backbone', 'sh

ows', 'aug', 'experts', 'combat', 'sharing', 'politician', 'investment', 'stakes', 'industry', 'servant', 'brave', 'seriously', 'somebody', 'set', 'join', 'doctor', 'vision', 'energy', 'week', 'promising', 'coast', 'dying', 'far', 'groups', 'grave', 'culture', '2018', 'people's', 'settle', 'getting', 'joining', 'fairly', 'public', 'however', 'unemployment', 'proudly', 'helped', 'alabama', 'grandkids', 'times', 'arrived', 'bully', 'difficult', 'vermont', 'disease', 'brutal', 'covid', 'encourage', 'realize', 'requires', 'nearly', 'marching', 'imagine', 'plasma', 'india', 'rebuilding', 'drive', 'r', 'catholic', 'dawn', 'fought', 'striving', 'deeply', 'destruction', 'foundation', 'changing', 'platform', 'determination', 'creed', 'university', 'pledged', 'peace', 'ever', 'david', 'entrepreneurs', 'feeling', 'honored', 'founded', '23', 'man', 'seeing', 'nominate', 'celebrated', 'money', 'michelle', 'nurse', 'taxpayer', 'straight', 'hand', 'beginning', 'short', 'constitutional', 'deaths', 'inspire', 'anthem', 'mental', 'statue', 'taylor', 'already', 'cell', 'let', 'double', 'you've', 'waters', '46', 'approach', 'cutting', 'paul', 'war', 'spanish', 'jail', 'white', 'ambition', 'closed', 'playing', 'claim', 'collapse', 'ballots', 'chose', 'past', 'later', 'racist', 'independence', 'swamp', 'scientists', 'divided', 'reelected', 'teaching', 'power', 'commitment', 'choices', 'phone', 'inspired', 'accept', 'event', 'caring', 'providing', 'fierce', 'taxes', 'comfort', 'g', 'baltimore', 'reached', 'go', 'justices', 'fight', 'fully', 'access', 'opportunities', 'fell', 'stops', 'thanks', 'navy', 'next', 'cuba', 'knowing', 'listen', 'faq', 'parents', 'asking', 'started', 'representative', 'lived', 'like', 'alive', 'rnc', 'accomplish', 'partisan', 'mobile', '9', 'womb', 'york', 'eliminating', 'safe', 'oval', 'insurance', 'guide', 'dictators', 'create', 'meeting', 'boom', 'introduced', 'possible', 'fantastic', 'jim', 'essential', '20', 'hispanic', 'inside', 'living', 'ended', 'though', 'nomination', 'isis', 'pick', 'write', 'politics', 'refused', 'creating', 'worked', 'elect', 'mark', 'defund', 'staff', 'breaking', 'escape', 'demand', 'professional', 'number', 'america's', 'haven't', 'reality', 'lifelong', 'trillion', 'opioid', 'background', 'n', 'rich', 'excited', 'carry', 'crazy', 'promoting', '30', 'affordable', 'land', 'announced', 'seek', 'fine', 'feet', 'breast', 'accomplished', 'teachers', 'protect', 'road', 'west', 'milwaukee', 'paying', 'questions', 'lead', 'pence', 'important', 'related'}



```
In [7]: def conv_features(text, fw) :
        """Given some text, this returns a dictionary holding the
           feature words.

           Args:
               * text: a piece of text in a continuous string. Assumes
               text has been cleaned and case folded.
               * fw: the *feature words* that we're considering. A word
               in `text` must be in fw in order to be returned. This
               prevents us from considering very rarely occurring words.

           Returns:
               A dictionary with the words in `text` that appear in `fw`.
               Words are only counted once.
               If `text` were "quick quick brown fox" and `fw` = {'quick','fox','jumps'},
               then this would return a dictionary of
               {'quick' : True,
                'fox' :    True}

        """

        # Your code here

        ret_dict = dict()

        words = text.split()
        for word in words:
            if word in fw:
                ret_dict[word] = True

        return(ret_dict)
```

```
In [8]: assert(len(feature_words)>0)

        assert(conv_features("donald is the president",feature_words)==
                {'donald':True,'president':True})
        assert(conv_features("people are american in america",feature_words)==
                {'america':True,'american':True,'people':True})
```

Now we'll build our feature set. Out of curiosity I did a train/test split to see how accurate the classifier was, but we don't strictly need to since this analysis is exploratory.

```
In [9]: featuresets = [(conv_features(text,feature_words), party) for (text, party) in conventi
```

```
In [10]: random.seed(20220507)
          random.shuffle(featuresets)

          test_size = 500
```

```
In [11]: test_set, train_set = featuresets[:test_size], featuresets[test_size:]
classifier = nltk.NaiveBayesClassifier.train(train_set)
print(nltk.classify.accuracy(classifier, test_set))
```

0.5

```
In [12]: classifier.show_most_informative_features(25)
```

Most Informative Features

china = True	Republ : Democr =	25.8 : 1.0
votes = True	Democr : Republ =	23.8 : 1.0
enforcement = True	Republ : Democr =	21.5 : 1.0
destroy = True	Republ : Democr =	19.2 : 1.0
freedoms = True	Republ : Democr =	18.2 : 1.0
climate = True	Democr : Republ =	17.8 : 1.0
supports = True	Republ : Democr =	17.1 : 1.0
crime = True	Republ : Democr =	16.1 : 1.0
media = True	Republ : Democr =	14.9 : 1.0
beliefs = True	Republ : Democr =	13.0 : 1.0
countries = True	Republ : Democr =	13.0 : 1.0
defense = True	Republ : Democr =	13.0 : 1.0
isis = True	Republ : Democr =	13.0 : 1.0
liberal = True	Republ : Democr =	13.0 : 1.0
religion = True	Republ : Democr =	13.0 : 1.0
trade = True	Republ : Democr =	12.7 : 1.0
flag = True	Republ : Democr =	12.1 : 1.0
greatness = True	Republ : Democr =	12.1 : 1.0
abraham = True	Republ : Democr =	11.9 : 1.0
defund = True	Republ : Democr =	11.9 : 1.0
drug = True	Republ : Democr =	10.9 : 1.0
department = True	Republ : Democr =	10.9 : 1.0
destroyed = True	Republ : Democr =	10.9 : 1.0
enemy = True	Republ : Democr =	10.9 : 1.0
amendment = True	Republ : Democr =	10.3 : 1.0

**My Observations:** NaiveBayes classify each party by each word in it, then give avg result. Over all it pretty much classifies the party with the words that a party normally would use.

## Part 2: Classifying Congressional Tweets

In this part we apply the classifier we just built to a set of tweets by people running for congress in 2018. These tweets are stored in the database congressional\_data.db. That DB is funky, so I'll give you the query I used to pull out the tweets. Note that this DB has some big tables and is unindexed, so the query takes a minute or two to run on my machine.

```
In [13]: congressional_data = sqlite3.connect("congressional_data.db")
cong_cur = congressional_data.cursor()
```

```
In [14]: results = cong_cur.execute(
    ...     SELECT DISTINCT
    ...         cd.candidate,
    ...         cd.party,
    ...         tw.tweet_text
    ...     FROM candidate_data cd
    ...     INNER JOIN tweets tw ON cd.twitter_handle = tw.handle
    ...     AND cd.candidate == tw.candidate
```

```

        AND cd.district == tw.district
        WHERE cd.party in ('Republican','Democratic')
        AND tw.tweet_text NOT LIKE '%RT%'
        '''

    )

results = list(results) # Just to store it, since the query is time consuming

```

```

In [15]: tweet_data = []

# Now fill up tweet_data with sublists like we did on the convention speeches.
# Note that this may take a bit of time, since we have a lot of tweets.
for row in results:
    # store the results in convention_data
    candidate, party, tweet_text = row
    tweet_text = str(tweet_text)
    tweet_text = str.lower(tweet_text)
    tweet_text = remove_punctuation(tweet_text)
    #text = tokenize(text)
    tweet_text = remove_stop(tweet_text)
    '''

    #apply(prepare, pipeline=my_pipeline)'''
    tweet_data.append((candidate, party, tweet_text))

```

```

In [16]: random.seed(20201014)

tweet_data_sample = random.choices(tweet_data,k=10)
print(tweet_data_sample)

```

```

[('Jimmy Panetta', 'Democratic', ' bearlier today spoke house floor abt protecting healt
h care women praised ppmarmonte work central coast httpstcowqgtrzt7vv'), ('Marcy Kaptu
r', 'Democratic', ' bgo tribe #rallytogether httpstco0nxutf1915'), ('Debbie Wasserman Sc
hultz', 'Democratic', ' bapparently trump thinks easy students overwhelmed crushing burd
en debt pay student loans #trumpbudget httpstcockyqo5t0qh'), ('Dave Brat', 'Republican',
' bwexe2x80x99re grateful first responders rescue personnel firefighters police voluntee
rs working tirelessly keep people safe provide muchneeded help putting lives linennhttps
tcoezpv0vmiz3'), ('Antonio Sabàto Jr', 'Republican', ' bletxe2x80x99s make even greater
#kag xf0x9fx87xbaxf0x9fx87xb8 httpstcoy9qozd5l2z'), ('Marcia Fudge', 'Democratic', ' bwe
1hr cavs tie series 22 im #allin216 repbarbaralee scared #roadtovictory'), ('Scott Peter
s', 'Democratic', ' bcongrats belliottd new gig sd city hall glad continue servexe2x80x
a6 httpstcofkvmw3cqdi'), ('Mariah Phillips', 'Democratic', ' bwe really close 3500 raise
d toward match right whoot thatxe2x80x99s 7000 nonmath majors room xf0x9fx98x82 help us
get httpstcotu34c472sd httpstcoqsdqkypsmc'), ('Jimmy Panetta', 'Democratic', ' btoday co
mment period potusxe2x80x99s plan expand offshore drilling opened public 60 days march 9
share oppose proposed program directly trump administration comments made email mail htt
pstcobaaymejxqn'), ('Lucille Roybal-Allard', 'Democratic', ' bcelebrated icseastlaxe2x80
x99s 22 years eastside commitment amp saluted community leaders last nightxe2x80x99s awa
rds dinner httpstco7v7gh8givb')]

```

```

In [17]: #creating fw

word_cutoff = 5

tokens = [w for c, p, t in tweet_data for w in t.split()]

word_dist = nltk.FreqDist(tokens)
feature_words = set()

```

```
for word, count in word_dist.items() :
    if count > word_cutoff :
        feature_words.add(word)
```

```
In [30]: featuresetss = [(party, conv_features(tweet_text, feature_words)) for (_, party, tweet_t
```

```
In [32]: for party, tweet in featuresetss :
    estimated_party = classifier.classify(tweet)
    # Fill in the right-hand side above with code that estimates the actual party

    print(f"Here's our (cleaned) tweet: {tweet}")
    print(f"Actual party is {party} and our classifier says {estimated_party}.")
    print("")
```

Here's our (cleaned) tweet: {'bearlier': True, 'today': True, 'spoke': True, 'house': True, 'floor': True, 'abt': True, 'protecting': True, 'health': True, 'care': True, 'women': True, 'praised': True, 'work': True, 'central': True, 'coast': True}

Actual party is Democratic and our classifier says Republican.

Here's our (cleaned) tweet: {'bgo': True, 'tribe': True, '#rallytogether': True}

Actual party is Democratic and our classifier says Democratic.

Here's our (cleaned) tweet: {'bapparently': True, 'trump': True, 'thinks': True, 'easy': True, 'students': True, 'overwhelmed': True, 'crushing': True, 'burden': True, 'debt': True, 'pay': True, 'student': True, 'loans': True, '#trumpbudget': True}

Actual party is Democratic and our classifier says Republican.

Here's our (cleaned) tweet: {'bwexe2x80x99re': True, 'grateful': True, 'first': True, 'responders': True, 'rescue': True, 'personnel': True, 'firefighters': True, 'police': True, 'volunteers': True, 'working': True, 'tirelessly': True, 'keep': True, 'people': True, 'safe': True, 'provide': True, 'muchneeded': True, 'help': True, 'putting': True, 'lives': True}

Actual party is Republican and our classifier says Republican.

Here's our (cleaned) tweet: {'bletxe2x80x99s': True, 'make': True, 'even': True, 'greater': True, '#kag': True, 'xf0x9fx87xbaxf0x9fx87xb8': True}

Actual party is Republican and our classifier says Republican.

Here's our (cleaned) tweet: {'bwe': True, 'cavs': True, 'tie': True, 'series': True, '22': True, 'im': True, '#allin216': True, 'repbarbaralee': True, 'scared': True, '#roadto victory': True}

Actual party is Democratic and our classifier says Republican.

Here's our (cleaned) tweet: {'bcongrats': True, 'new': True, 'gig': True, 'sd': True, 'city': True, 'hall': True, 'glad': True, 'continue': True}

Actual party is Democratic and our classifier says Republican.

Here's our (cleaned) tweet: {'bwe': True, 'really': True, 'close': True, '3500': True, 'raised': True, 'toward': True, 'match': True, 'right': True, 'thatxe2x80x99s': True, '7000': True, 'majors': True, 'room': True, 'xf0x9fx98x82': True, 'help': True, 'us': True, 'get': True}

Actual party is Democratic and our classifier says Republican.

Here's our (cleaned) tweet: {'btoday': True, 'comment': True, 'period': True, 'potusxe2x80x99s': True, 'plan': True, 'expand': True, 'offshore': True, 'drilling': True, 'opened': True, 'public': True, '60': True, 'days': True, 'march': True, '9': True, 'share': True, 'oppose': True, 'proposed': True, 'program': True, 'directly': True, 'trump': True, 'administration': True, 'comments': True, 'made': True, 'email': True, 'mail': True}

Actual party is Democratic and our classifier says Republican.

Here's our (cleaned) tweet: {'bcelebrated': True, '22': True, 'years': True, 'eastside':

True, 'commitment': True, 'amp': True, 'saluted': True, 'community': True, 'leaders': True, 'last': True, 'nightxe2x80x99s': True, 'awards': True, 'dinner': True}  
 Actual party is Democratic and our classifier says Republican.

```
In [38]: # dictionary of counts by actual party and estimated party.
# first key is actual, second is estimated
parties = ['Republican', 'Democratic']
results = defaultdict(lambda: defaultdict(int))

for p in parties :
    for p1 in parties :
        results[p][p1] = 0

num_to_score = 10000
random.shuffle(tweet_data)

featuresetss = [(party, conv_features(tweet_text, feature_words)) for (_, party, tweet_

for idx, tp in enumerate(featuresetss) :
    party, tweet = tp
    # Now do the same thing as above, but we store the results rather
    # than printing them.

    # get the estimated party
    estimated_party = classifier.classify(tweet)

    results[party][estimated_party] += 1

    if idx > num_to_score :
        break
```

```
In [39]: results
```

```
Out[39]: defaultdict(<function __main__.<lambda>()>,
    {'Republican': defaultdict(int,
        {'Republican': 3646, 'Democratic': 736}),
    'Democratic': defaultdict(int,
        {'Republican': 4551, 'Democratic': 1069})})
```

## Reflections

NaiveBayes predict very accurate on Republican, but a little wrong with Democratic class, maybe because our data is not balanced