Libraries

We need four main libraries for this lecture: Scikit learn, numpy, scipy, and Pandas. An easy way to install all of them in one go is to get Anaconda: https://anaconda.org/anaconda/python. When you install Anaconda you will also get access to Jupyter notebook using which you can run all the below commands. An alternative, which I highly recommend is to first install Anaconda and run your code using editors such as Text Wrangle, Sublime Text, Notepad ++, or Vim editor. Doing the later will highly be beneficial for you after this course is over (but it does have a learning curve).

Data

For this practice class lets take the data from the first assignment.

Using our libraries to import the data

In [321]:
```
import pandas as pd
import numpy as np
import scipy
```

In [322]:
```
print("Hello world!")
```

```
Hello world!
```

In [41]:
```
df = pd.read_csv("Location_of_you_folder/train.csv",encoding='ISO-8859-1')
 # there are different encoding and for HW1 this particular encoding works
 well.
```

In [327]:
```
df.head(2)
```

Out[327]:

	text	class
0	It was clear right from the beginning that 9/...	positive
1	The most hillarious and funny Brooks movie I ...	positive

In [328]:
```
df.columns ## To print out the column names
```

Out[328]:
```
Index(['text', 'class'], dtype='object')
```

In [325]:
```
df.shape ## To know the dimensions of the data frame. The output should be
 read as [rows,columns]
```

Out[325]:
```
(2000, 2)
```

In [23]:
```
[rows,columns] = df.shape # Print both values out to check the output your self.
```

In [329]:
```
df.describe()
```

Out[329]:

	text	class
count	2000	2000
unique	2000	2
top	My title above says it all. Let me make it cl...	negative
freq	1	1000

In [339]:
```
df['class'].unique() ## To know the unique labels.
```

Out[339]:
```
array(['positive', 'negative'], dtype=object)
```

In [331]:
```
df['class'].value_counts() ## To know the distribution of the labels.
```

Out[331]:
```
negative    1000
positive    1000
Name: class, dtype: int64
```

Convert labels into a machine readable format

In [332]:
```
from sklearn import preprocessing ### Importing a preprocessor to convert the labels in the target class.
```

In [333]:
```
train_class_y = ['negative','positive']
```

In [334]:
```
le = preprocessing.LabelEncoder() ## Label encoder does the trick.
```

In [335]:
```
le.fit(train_class_y) ## We are fitting the categories now.
```

Out[335]:
```
LabelEncoder()
```

In [337]:
```
train_y = le.transform(df['class'])
```

In [340]:
```
train_y ## These are out labels now.
```

Out[340]:
```
array([1, 1, 1, ..., 0, 0, 0])
```

In [341]:
```
le.transform(['positive','negative','positive']) ### Just to check
```

Out[341]:
```
array([1, 0, 1])
```

In [342]:
```
le.inverse_transform([0,1,1]) ### Doing an inverse of the transformation
```

Out[342]:
```
array(['negative', 'positive', 'positive'],
      dtype='<U8')
```

Tackling our text data

In [343]:
```
train_x = df['text']
```

In [344]:
```
train_x.shape ## Just making sure that we have what we want.
```

Out[344]:
```
(2000,)
```

Rather than doing bag of words, we are goint to convert our data into tf.idf's

In [345]:
```
from sklearn.feature_extraction.text import TfidfVectorizer ## Importing t
he library that will help us do this.
```

In [424]:
```
tf = TfidfVectorizer(min_df=1,stop_words='english',max_features=5000)  ## A
sk yourself, why min_df =1? We are using english stopwords.
####max_features=3000
```

In [425]:
```
train_x_tfidf = tf.fit_transform(train_x)
```

In [426]:
```
tf.get_feature_names()  ## Be careful to check your feature names with tf a
nd not with train_x_tfidf
```

Out[426]:
```
['000',
 '10',
 '100',
 '101',
 '11',
 '12',
 '13',
 '14',
 '15',
 '16',
 '17',
 '18',
 '1930',
 '1932',
 '1933',
 '1939',
 '1943',
 '1950',
 '1950s',
 '1960s',
 '1964',
 '1965',
 '1970',
 '1971',
 '1973',
 '1978',
 '1979',
 '1980',
 '1982',
```

```
                 '1983',
                 '1984',
                 '1990',
                 '1995',
                 '1998',
                 '1999',
                 '1st',
                 '20',
                 '2000',
                 '2001',
                 '2002',
                 '2003',
                 '2004',
                 '2005',
                 '2006',
                 '2007',
                 '20s',
                 '20th',
                 '21st',
                 '24',
                 '25',
                 '26',
                 '28',
                 '2nd',
                 '30',
                 '30s',
                 '35',
                 '3d',
                 '3rd',
                 '40',
                 '45',
                 '50',
                 '60',
                 '60s',
                 '70',
                 '70s',
                 '747',
                 '77',
                 '80',
                 '80s',
                 '90',
                 '90s',
                 '95',
                 '99',
                 'abandoned',
                 'abc',
                 'abilities',
                 'ability',
                 'able',
                 'abound',
                 'abraham',
                 'absence',
                 'absent',
                 'absolute',
                 'absolutely',
                 'absurd',
                 'absurdity',
```

```
 'abu',
 'abundance',
 'abuse',
 'abusive',
 'abysmal',
 'academy',
 'accent',
 'accents',
 'accept',
 'acceptable',
 'accepted',
 'access',
 'accident',
 'accidentally',
 'acclaimed',
 'accompanied',
 'accomplished',
 'according',
 'account',
 'accounts',
 'accuracy',
 'accurate',
 'accurately',
 'accused',
 'achieve',
 'achieved',
 'achievement',
 'acid',
 'act',
 'acted',
 'acting',
 'action',
 'actions',
 'active',
 'actor',
 'actors',
 'actress',
 'actresses',
 'acts',
 'actual',
 'actually',
 'ad',
 'adam',
 'adams',
 'adaptation',
 'adaptations',
 'adapted',
 'adaption',
 'add',
 'added',
 'addict',
 'adding',
 'addition',
 'addressed',
 'adds',
 'adequate',
 'admit',
```

```
                 'admittedly',
                 'adopted',
                 'adorable',
                 'adult',
                 'adults',
                 'advance',
                 'advantage',
                 'adventure',
                 'adventures',
                 'advertised',
                 'advertising',
                 'advice',
                 'affair',
                 'affect',
                 'affected',
                 'affection',
                 'aforementioned',
                 'afraid',
                 'africa',
                 'african',
                 'afternoon',
                 'age',
                 'aged',
                 'agent',
                 'agents',
                 'ages',
                 'aggressive',
                 'aging',
                 'ago',
                 'agony',
                 'agree',
                 'ah',
                 'ahead',
                 'aid',
                 'aim',
                 'aimed',
                 'aiming',
                 'ain',
                 'air',
                 'aired',
                 'airport',
                 'ajax',
                 'aka',
                 'akshay',
                 'al',
                 'alan',
                 'alas',
                 'albeit',
                 'albert',
                 'album',
                 'alcohol',
                 'alcoholic',
                 'alert',
                 'alex',
                 'alexandre',
                 'alfred',
                 'alice',
```

```
                 'alien',
                 'aliens',
                 'alike',
                 'alive',
                 'alleged',
                 'allen',
                 'alligator',
                 'allow',
                 'allowed',
                 'allowing',
                 'allows',
                 'alongside',
                 'alright',
                 'altered',
                 'alternative',
                 'altman',
                 'altogether',
                 'amateur',
                 'amateurish',
                 'amazed',
                 'amazing',
                 'amazingly',
                 'ambiguous',
                 'ambitious',
                 'america',
                 'american',
                 'americans',
                 'amusing',
                 'analysis',
                 'ancient',
                 'anderson',
                 'andrew',
                 'andy',
                 'angel',
                 'angeles',
                 'angels',
                 'anger',
                 'angles',
                 'angry',
                 'animal',
                 'animals',
                 'animated',
                 'animation',
                 'anime',
                 'ann',
                 'anna',
                 'anne',
                 'annie',
                 'annoyed',
                 'annoying',
                 'answer',
                 'answers',
                 'anthony',
                 'anti',
                 'antics',
                 'anton',
                 'antonioni',
```

```
'anybody',
'anymore',
'anytime',
'anyways',
'apart',
'apartment',
'ape',
'apes',
'appalling',
'apparent',
'apparently',
'appeal',
'appealing',
'appear',
'appearance',
'appearances',
'appeared',
'appearing',
'appears',
'appreciate',
'appreciation',
'approach',
'appropriate',
'april',
'area',
'areas',
'aren',
'arguably',
'argument',
'arkin',
'arms',
'army',
'arnie',
'arnold',
'arquette',
'arrested',
'arrival',
'arrives',
'arriving',
'arrogance',
'arrogant',
'art',
'arthur',
'artist',
'artistic',
'artists',
'arts',
'artsy',
'artwork',
'ashamed',
'asian',
'aside',
'ask',
'asked',
'asking',
'asks',
'asleep',
```

```
                  'aspect',
                  'aspects',
                  'ass',
                  'assassin',
                  'assigned',
                  'associated',
                  'assume',
                  'assumed',
                  'astonishing',
                  'asylum',
                  'atlantis',
                  'atmosphere',
                  'atmospheric',
                  'atrocious',
                  'attached',
                  'attack',
                  'attacked',
                  'attacks',
                  'attempt',
                  'attempted',
                  'attempting',
                  'attempts',
                  'attenborough',
                  'attend',
                  'attention',
                  'attila',
                  'attitude',
                  'attitudes',
                  'attracted',
                  'attraction',
                  'attractive',
                  'audience',
                  'audiences',
                  'audio',
                  'august',
                  'aunt',
                  'aussie',
                  'austen',
                  'austin',
                  'australia',
                  'australian',
                  'australians',
                  'auteur',
                  'authentic',
                  'authenticity',
                  'author',
                  'authorities',
                  'authority',
                  'automatically',
                  'available',
                  'average',
                  'avoid',
                  'awake',
                  'awakening',
                  'award',
                  'awards',
                  'aware',
```

```
'away',
'awe',
'awesome',
'awful',
'awhile',
'awkward',
'baby',
'bacall',
'backdrop',
'backed',
'background',
'backgrounds',
'backs',
'bad',
'badly',
'bag',
'bakshi',
'balance',
'balduin',
'ball',
'balls',
'bamboo',
'banana',
'band',
'bands',
'bang',
'bank',
'banned',
'bar',
'barbara',
'barbra',
'bare',
'barely',
'bargain',
'barman',
'barnes',
'baron',
'barry',
'bart',
'base',
'baseball',
'based',
'basement',
'bashing',
'basic',
'basically',
'basinger',
'basis',
'basketball',
'bath',
'batman',
'battle',
'battles',
'bbc',
'beach',
'bear',
'bears',
```

```
                 'beast',
                 'beat',
                 'beaten',
                 'beating',
                 'beatty',
                 'beautiful',
                 'beautifully',
                 'beauty',
                 'bed',
                 'beer',
                 'began',
                 'begin',
                 'beginning',
                 'begins',
                 'behavior',
                 'behaviour',
                 'behold',
                 'beings',
                 'bela',
                 'belief',
                 'beliefs',
                 'believable',
                 'believe',
                 'believed',
                 'believes',
                 'believing',
                 'belongs',
                 'beloved',
                 'ben',
                 'benefit',
                 'berlin',
                 'bernard',
                 'best',
                 'bet',
                 'bette',
                 'better',
                 'bettie',
                 'beverly',
                 'beware',
                 'biased',
                 'bible',
                 'big',
                 'bigger',
                 'biggest',
                 'biko',
                 'bilko',
                 'billed',
                 'billy',
                 'bin',
                 'biography',
                 'birds',
                 'birth',
                 'birthday',
                 'bit',
                 'bite',
                 'bits',
                 'bitten',
```

```
                'bitter',
                'bizarre',
                'black',
                'blade',
                'blah',
                'blair',
                'blake',
                'blame',
                'bland',
                'blank',
                'blanks',
                'blatant',
                'bleak',
                'bleed',
                'blew',
                'blind',
                'block',
                'blockbuster',
                'blonde',
                'blood',
                'bloodshed',
                'bloody',
                'bloom',
                'blow',
                'blowing',
                'blown',
                'blows',
                'blue',
                'blunt',
                'board',
                'boat',
                'bob',
                'bobby',
                'bodies',
                'body',
                'bold',
                'boll',
                'bollywood',
                'bomb',
                'bomber',
                'bombshells',
                'bonanza',
                'bond',
                'bone',
                'bonham',
                'bonus',
                'boogeyman',
                'book',
                'books',
                'boom',
                'border',
                'bore',
                'bored',
                'boredom',
                'boring',
                'born',
                'boss',
```

```
                 'bother',
                 'bothered',
                 'bouchet',
                 'bought',
                 'bound',
                 'bourne',
                 'bouzaglo',
                 'bowl',
                 'box',
                 'boxer',
                 'boxing',
                 'boy',
                 'boyfriend',
                 'boyle',
                 'boys',
                 'brad',
                 'brady',
                 'brain',
                 'brains',
                 'branagh',
                 'brand',
                 'brave',
                 'break',
                 'breaking',
                 'breaks',
                 'breasts',
                 'breath',
                 'breathtaking',
                 'brian',
                 'bride',
                 'bridge',
                 'brief',
                 'briefly',
                 'bright',
                 'brilliant',
                 'brilliantly',
                 'bring',
                 'bringing',
                 'brings',
                 'britain',
                 'british',
                 'broad',
                 'broadcast',
                 'broadway',
                 'broke',
                 'broken',
                 'bronte',
                 'brooding',
                 'brooklyn',
                 'brooks',
                 'brosnan',
                 'brother',
                 'brothers',
                 'brought',
                 'brown',
                 'bruce',
                 'brutal',
```

```
                 'brutally',
                 'bsg',
                 'buck',
                 'bucks',
                 'buddies',
                 'buddy',
                 'budget',
                 'buffs',
                 'bugs',
                 'build',
                 'building',
                 'buildings',
                 'builds',
                 'built',
                 'bulk',
                 'bullet',
                 'bullets',
                 'bully',
                 'bumbling',
                 'bunch',
                 'buried',
                 'burke',
                 'burn',
                 'burned',
                 'burning',
                 'burns',
                 'burton',
                 'bus',
                 'bush',
                 'business',
                 'businessman',
                 'busy',
                 'butcher',
                 'butt',
                 'buy',
                 'buying',
                 'cabin',
                 'cable',
                 'cage',
                 'cagney',
                 'cain',
                 'caine',
                 'cake',
                 'caliber',
                 'california',
                 'called',
                 'calling',
                 'calls',
                 'came',
                 'cameo',
                 'cameos',
                 'camera',
                 'cameras',
                 'camp',
                 'campaign',
                 'campbell',
                 'campy',
```

```
'canada',
'canadian',
'cancer',
'candy',
'cannon',
'canto',
'capable',
'capital',
'capshaw',
'captain',
'captivating',
'capture',
'captured',
'captures',
'capturing',
'car',
'card',
'cards',
'care',
'cared',
'career',
'careers',
'careful',
'carefully',
'cares',
'caring',
'carol',
'carpenter',
'carradine',
'carrie',
'carried',
'carries',
'carry',
'carrying',
'cars',
'carter',
'cartoon',
'cartoonish',
'cartoons',
'caruso',
'case',
'cases',
'cash',
'cassidy',
'cast',
'casting',
'castle',
'casual',
'cat',
'catch',
'catching',
'catchy',
'category',
'catherine',
'catholic',
'caucasian',
'caught',
```

```
'cause',
'caused',
'cave',
'cd',
'celine',
'cell',
'cells',
'celluloid',
'cena',
'center',
'centered',
'centers',
'central',
'cents',
'century',
'certain',
'certainly',
'cg',
'cgi',
'chain',
'chair',
'challenge',
'challenged',
'champion',
'championship',
'chan',
'chance',
'chances',
'change',
'changed',
'changes',
'changing',
'channel',
'chaos',
'character',
'characterization',
'characters',
'charge',
'charisma',
'charismatic',
'charles',
'charlie',
'charlotte',
'charm',
'charming',
'chase',
'chased',
'chasing',
'chavez',
'che',
'cheap',
'cheaply',
'cheating',
'check',
'checked',
'checking',
'cheech',
```

```
'cheek',
'cheese',
'cheesy',
'chemistry',
'chick',
'chief',
'child',
'childhood',
'children',
'chill',
'chilling',
'chills',
'china',
'chinese',
'choice',
'choices',
'choir',
'chomsky',
'chong',
'choose',
'chooses',
'choreographed',
'choreography',
'chorus',
'chose',
'chosen',
'chris',
'chrissy',
'christ',
'christian',
'christmas',
'christopher',
'chuckle',
'church',
'cia',
'cinderella',
'cinema',
'cinematic',
'cinematography',
'circumstances',
'citizen',
'citizens',
'city',
'civil',
'civilization',
'claim',
'claimed',
'claiming',
'claims',
'claire',
'clarity',
'clark',
'clarke',
'clash',
'class',
'classic',
'classics',
```

```
'claude',
'clean',
'clear',
'clearly',
'clever',
'cliche',
'clichã',
'cliff',
'cliffhanger',
'climactic',
'climax',
'clip',
'clips',
'cloak',
'clock',
'clooney',
'close',
'closer',
'closest',
'closing',
'clothes',
'clothing',
'clown',
'club',
'clue',
'clueless',
'clues',
'clumsy',
'coach',
'coast',
'coaster',
'code',
'cohen',
'coherent',
'cold',
'cole',
'collar',
'collection',
'collective',
'collector',
'college',
'collette',
'colman',
'colonel',
'color',
'colorful',
'colors',
'colour',
'colours',
'columbo',
'com',
'combat',
'combination',
'combine',
'combined',
'combines',
'come',
```

```
                 'comedian',
                 'comedians',
                 'comedic',
                 'comedies',
                 'comedy',
                 'comes',
                 'comfortable',
                 'comic',
                 'comical',
                 'coming',
                 'command',
                 'commendable',
                 'comment',
                 'commentary',
                 'commented',
                 'comments',
                 'commercial',
                 'commercials',
                 'commit',
                 'commitment',
                 'committed',
                 'common',
                 'communist',
                 'community',
                 'company',
                 'compare',
                 'compared',
                 'comparing',
                 'comparison',
                 'compelled',
                 'compelling',
                 'competent',
                 'competition',
                 'complain',
                 'complaining',
                 'complaint',
                 'complete',
                 'completely',
                 'complex',
                 'complicated',
                 'composed',
                 'composer',
                 'compromise',
                 'computer',
                 'concept',
                 'concerned',
                 'concerning',
                 'concerns',
                 'concert',
                 'conclusion',
                 'condemned',
                 'condition',
                 'confess',
                 'confession',
                 'conflict',
                 'confused',
                 'confusing',
```

```
                  'confusion',
                  'connect',
                  'connected',
                  'connection',
                  'connections',
                  'connery',
                  'connolly',
                  'connor',
                  'conquest',
                  'conservative',
                  'consider',
                  'considerable',
                  'considered',
                  'considering',
                  'consistent',
                  'consistently',
                  'consists',
                  'conspiracy',
                  'constant',
                  'constantly',
                  'construction',
                  'contact',
                  'contain',
                  'contained',
                  'contains',
                  'contemporary',
                  'content',
                  'contest',
                  'contestants',
                  'context',
                  'continue',
                  'continued',
                  'continues',
                  'continuing',
                  'continuity',
                  'contract',
                  'contrary',
                  'contrast',
                  'contribution',
                  'contrived',
                  'control',
                  'controlled',
                  'controversial',
                  'conversation',
                  'convey',
                  'conveyed',
                  'conviction',
                  'convince',
                  'convinced',
                  'convinces',
                  'convincing',
                  'convincingly',
                  'convoluted',
                  'cook',
                  'cooking',
                  'cool',
                  'cooper',
```

```
 'cop',
 'cope',
 ...]
```

In [427]:
```
train_x_tfidf_array = train_x_tfidf.toarray()
```

In [428]:
```
train_x_tfidf_array[0]
```

Out[428]:
```
array([ 0.,  0.,  0., ...,  0.,  0.,  0.])
```

In [429]:
```
tf.inverse_transform(train_x_tfidf_array[0]) ## just to check what all fea
tures are there.
```

Out[429]:
```
[array(['11', '1973', 'add', 'air', 'alas', 'angry', 'anti', 'appealing',
        'appearing', 'audience', 'bad', 'beginning', 'best', 'big', 'bin',
        'bizarre', 'black', 'bomb', 'boys', 'bunch', 'calls', 'car',
        'certainly', 'cheesy', 'choreography', 'cia', 'cinema', 'classic',
        'claude', 'clear', 'collective', 'combined', 'come', 'consider',
        'cuts', 'danger', 'deaf', 'did', 'different', 'directly',
        'director', 'directors', 'don', 'double', 'easy', 'effort',
        'elected', 'end', 'ending', 'entire', 'episodes', 'ernest', 'event
',
        'example', 'explain', 'extremely', 'faces', 'falling', 'family',
        'far', 'features', 'film', 'films', 'finally', 'forget', 'girl',
        'good', 'great', 'happened', 'happy', 'hard', 'haven', 'help',
        'henry', 'hope', 'hysterical', 'ii', 'imagine', 'impressive',
        'indian', 'inspire', 'instead', 'interesting', 'international',
        'just', 'ken', 'killed', 'known', 'leading', 'life', 'll', 'long',
        'love', 'luck', 'mainly', 'make', 'man', 'masterpiece', 'mentioned
',
        'minute', 'movies', 'muslim', 'needless', 'new', 'open', 'opening'
,
        'order', 'parts', 'pearl', 'penn', 'people', 'phone', 'pictures',
        'portrayal', 'president', 'pretty', 'probably', 'prologue',
        'promise', 'question', 'react', 'real', 'really', 'recognition',
        'recognize', 'release', 'remembered', 'reporter', 'reward', 'right
',
        'sad', 'say', 'screen', 'sean', 'segment', 'segments', 'september'
,
        'shares', 'shocked', 'silence', 'son', 'sound', 'starring',
        'starting', 'story', 'strange', 'suggests', 'sure', 'surprisingly'
,
        'takes', 'tale', 'talking', 'tastes', 'terrible', 'things',
        'thirty', 'tower', 'towers', 'tradition', 'tries', 'trying', 'twin
',
        'unique', 'vietnam', 'war', 'watch', 'way', 'western', 'women',
        'work', 'world', 'years', 'yes', 'york'],
       dtype='<U16')]
```

Importing a learning model

Let us start with a Multinomial Naive Bayes

In [354]:
```
from sklearn.naive_bayes import MultinomialNB
```

In [471]:
```
mnb = MultinomialNB(alpha=1.0) # Check what this alpha value is. You have already learnt most of the math to understand this.
```

In [376]:
```
mnb.fit(train_x_tfidf_array,train_y)
```

Out[376]:
```
MultinomialNB(alpha=1.0, class_prior=None, fit_prior=True)
```

Lets get out test data now

In [430]:
```
test_df = pd.read_csv("/Users/boY/Desktop/inls613_fall/class_exercise_Data/test.csv",encoding='ISO-8859-1')
```

In [431]:
```
test_x_tfidf = tf.transform(test_df['text'])
```

In [432]:
```
test_x_tfidf_array = test_x_tfidf.toarray()
```

In [433]:
```
test_y = le.transform(test_df['class'])
```

In [434]:
```
test_y.shape
```

Out[434]:
```
(2000,)
```

In [435]:
```
test_x_tfidf_array.shape
```

Out[435]:
```
(2000, 5000)
```

In [382]:
```
predictions = mnb.predict(test_x_tfidf_array)
```

In [383]:
```
predictions.shape
```

Out[383]:
```
(2000,)
```

In [384]:
```
count = 0
for i in range (len(predictions)):
    if predictions[i]==test_y[i]:
        count=count+1
```

In [385]:
```
count
```

Out[385]:
```
1651
```

In [386]:
```
count/2000
```

Out[386]:
```
0.8255
```

Feature engineering: Pick the top tf idf features

In [114]:
```
feature_array = np.array(tf.get_feature_names())
```

In [115]:
```
tfidf_sorting = np.argsort(train_x_tfidf.toarray()).flatten()[::-1]
```

In [116]:
```
tfidf_sorting
```

Out[116]:
```
array([17528, 25411, 11562, ..., 16921, 16922,     0])
```

In [117]:
```
n = 500
```

In [126]:
```
top_n = feature_array[tfidf_sorting][:n]
```

In [127]:
```
top_n
```

Out[127]:
```
array(['prison', 'zombie', 'infected', 'prisoners', 'guy', 'scientist',
        'zombies', 'military', 'da', 'mob', 'hero', 'serum', 'bars',
        'journalist', 'prisoner', 'jail', 'cure', 'scene', 'boss', 'goes',
        'way', 'long', 'deal', 'save', 'rasta', 'investigative', 'steadies'
,
        'electrocution', 'communion', 'gracing', 'badie', 'empowered',
        'mobs', 'scrambles', 'kids', 'awfulness', 'guards', 'takes',
        'secure', 'irate', 'instruction', 'decapitation', 'tendencies',
        'swat', 'munching', 'woman', 'main', 'make', 'pressed', 'shreds',
        'experimenting', 'hallway', 'rapes', 'manual', 'crashes', 'babe',
        'freshly', 'vet', 'ripping', 'researching', 'warden', 'imprisoned',
        'worth', 'gets', 'invented', 'psychotic', 'charge', 'riot', 'wine',
        'teams', 'homicidal', 'gangs', 'route', 'guard', 'noting',
        'position', 'bodies', 'corrupt', 'path', 'gross', 'terror',
        'highlight', 'real', 'movie', 'vietnam', 'practically', 'scenes',
        '80s', 'gold', 'blow', 'hundreds', 'plans', 'criminal', 'ultra',
        'scare', 'calls', 'bits', 'round', 'ready', 'reach', 'gory', 'ex',
        'behavior', 'involving', 'pulled', 'ground', 'bloody', 'soldiers',
        'flicks', '40', 'breaks', 'faces', 'gem', 'super', 'independent',
        'pass', 'wall', 'appearance', 'hey', 'inside', 'disturbing',
        'genius', 'contains', 'spoilers', 'caught', 'hair', 'crazy',
        'escape', 'outside', 'creepy', 'cheesy', 'film', 'team', 'plain',
        'slightly', 'create', 'clichã', 'lady', 'nearly', 'hot', 'hands',
        'leads', 'sets', 'killed', 'heart', 'brother', 'genre', 'kill',
        'turned', 'wants', 'head', 'overall', 'problem', 'early', 'moments'
,
        'fans', 'version', 'gives', 'dead', 'classic', 'use', 'couple',
        'half', 'mind', 'place', 'different', 'course', 'sense', 'minutes',
        'original', 'trying', 'horror', 'making', 'want', 'new', 'look',
        'lot', 'going', 'watching', 'end', 'character', 'plot', 'people',
        'time', 'good', 'just', 'finest', 'feisty', 'feelings', 'fitting',
        'fittingly', 'feldshuh', 'fits', 'feline', 'felix', 'fitted',
        'feinstone', 'feinting', 'finger', 'feeds', 'finneran', 'feijã³',
        'feigned', 'feifel', 'fell', 'fitzgerald', 'fees', 'feels', 'feel',
        'feelgood', 'feeling', 'feet', 'fitfully', 'fella', 'fend', 'femme'
,
        'fisted', 'fence', 'fine', 'finney', 'fencer', 'fencing', 'fist',
        'fellas', 'fends', 'finds', 'fenton', 'fenway', 'findings',
        'fishing', 'fishermen', 'feminists', 'finely', 'feminist',
        'femininity', 'fellini', 'felliniesque', 'feedback', 'fellow',
        'fit', 'fistsof', 'fellows', 'fellowship', 'fisticuffs', 'felon',
```

```
        'felons', 'felt', 'female', 'females', 'feminine', 'feeding',
        'filmfour', 'feed', 'fitzpatrick', 'faze', 'faã', 'fbi', 'flag',
        'fc', 'fdr', 'flabby', 'fear', 'finished', 'flabbergasted',
        'feared', 'fearful', 'fearing', 'finish', 'fearless', 'fears',
        'finis', 'fay', 'finishes', 'fawcett', 'favorite', 'faust', 'faux',
        'fav', 'flagrantly', 'faves', 'favor', 'favored', 'favorites',
        'favours', 'favors', 'favour', 'finishing', 'flagging', 'favoured',
        'favourite', 'favourites', 'fearsome', 'feasibility', 'feast',
        'fedja', 'february', 'fecundity', 'fixate', 'fed', 'fingered',
        'federal', 'federation', 'fix', 'feb', 'fedor8', 'feds', 'finlay',
        'ferdinand', 'feeb', 'feeble', 'fiver', 'fingering', 'featuring',
        'fingertips', 'feats', 'fl', 'fingers', 'fizzled', 'feat',
        'feather', 'feathers', 'fingernails', 'fixing', 'featurettes',
        'feature', 'featured', 'fixer', 'features', 'featurette', 'fixed',
        'fixated', 'ferch', 'financing', 'ferhan', 'firstly', 'fish',
        'fischer', 'fiennes', 'fierce', 'firth', 'fiery', 'firefighter',
        'firmly', 'fiend', 'fife', 'firefighters', 'fifi', 'fifth',
        'filmography', 'fifties', 'fight', 'fiendish', 'fields', 'fighter',
        'fictionalized', 'fickle', 'fiction', 'fictional', 'fishburne',
        'fictionalising', 'fishbourne', 'filter', 'fictitious', 'fielder',
        'fiddle', 'filone', 'fidel', 'films', 'fidelity', 'fido', 'field',
        'firm', 'fighters', 'feriss', 'fills', 'fireman', 'filipino',
        'filipinosâ', 'filmmakers', 'filled', 'filler', 'filling',
        'filmmaker', 'filed', 'filmable', 'filmcow', 'filmcritics',
        'filmdom', 'filming', 'filmed', 'filmic', 'files', 'file',
        'fightin', 'firguring', 'fighting', 'fightm', 'fights',
        'figuratively', 'figure', 'firing', 'filmmmakers', 'figured', 'fil'
,
        'fireworks', 'figurehead', 'fireplace', 'figures', 'figurines',
        'filmmaking', 'figuring', 'filtering', 'ficker', 'faultâ', 'fervor'
,
        'fisherman', 'ferries', 'ferris', 'ferry', 'finances', 'ferryman',
        'fertile', 'fessenden', 'ferrell', 'fest', 'festering', 'fisher',
        'festival', 'finance', 'festivals', 'festive', 'ferrer', 'ferraris'
,
        'fibre', 'fernack', 'finding', 'ferment', 'finders', 'finch',
        'firehouse', 'fermi', 'financially', 'fernandes', 'ferrari',
        'fernando', 'ferocious', 'financial', 'finnish', 'feroze',
        'ferrara', 'firearms', 'festivism', 'festivities', 'finland', 'fez'
,
        'fim', 'filthy', 'feuding', 'feudâ', 'firebombs', 'fever', 'fewer',
        'ff2', 'finally', 'fi', 'fired', 'filthiest', 'filth', 'fi9lm',
        'fiance', 'fiancã', 'feudal', 'feud', 'fetuses', 'fetus', 'fetch',
        'fetched', 'fetching', 'finality', 'firebird', 'fetchingly',
        'fetchit', 'fetid', 'fetish', 'fetishes', 'finale', 'fetishistic',
        'fetishwear', 'final', 'fettle', 'fichtner', 'â½', 'faulty',
        'explosives', 'exponential', 'exports', 'expose', 'exposed',
        'exposer', 'exposes'],
       dtype='<U30')
```

Different classification algorithm: Logistic Regression

In [436]:
```
from sklearn.linear_model import LogisticRegression # load the library
```

In [443]:
```
log_reg = LogisticRegression(C=4.0)
```

In [444]:
```
log_reg.fit(train_x_tfidf_array,train_y)
```

Out[444]:
```
LogisticRegression(C=4.0, class_weight=None, dual=False, fit_intercept=Tru
e,
          intercept_scaling=1, max_iter=100, multi_class='ovr', n_jobs=1,
          penalty='l2', random_state=None, solver='liblinear', tol=0.0001,
          verbose=0, warm_start=False)
```

In [445]:
```
log_reg.score(train_x_tfidf_array,train_y) # running it on the train set i
tself.
```

Out[445]:
```
0.99050000000000005
```

In [447]:
```
log_reg.score(test_x_tfidf_array,test_y) # running it on the test set.
```

Out[447]:
```
0.84099999999999997
```

Different classification algorithm: SVM

In [448]:
```
from sklearn import svm
```

In [449]:
```
clf = svm.SVC(C=2.0,degree=1,kernel='linear')
```

In [450]:
```
clf.fit(train_x_tfidf_array,train_y)
```

Out[450]:
```
SVC(C=2.0, cache_size=200, class_weight=None, coef0=0.0,
  decision_function_shape='ovr', degree=1, gamma='auto', kernel='linear',
  max_iter=-1, probability=False, random_state=None, shrinking=True,
  tol=0.001, verbose=False)
```

In [451]:
```
predicted = clf.predict(test_x_tfidf_array)
```

In [452]:
```
count = 0
for i in range (len(predicted)):
    if predicted[i]==test_y[i]:
        count=count+1
count
```

Out[452]:
```
1636
```

In [453]:
```
len(predicted)
```

Out[453]:
```
2000
```

In [472]:
```
1636/2000 # Computing accurarcy here.
```

Out[472]:
```
0.818
```

Different classification algorithm: Random Forest

In [455]:
```
from sklearn.ensemble import RandomForestClassifier
```

In [463]:
```
forest = RandomForestClassifier(max_depth=10,n_estimators=100,min_samples_
leaf=2)
#max_depth=10,n_estimators=100,min_samples_leaf=2
```

In [464]:
```
forest.fit(train_x_tfidf_array,train_y)
```

Out[464]:
```
RandomForestClassifier(bootstrap=True, class_weight=None, criterion='gini'
,
            max_depth=10, max_features='auto', max_leaf_nodes=None,
            min_impurity_decrease=0.0, min_impurity_split=None,
            min_samples_leaf=2, min_samples_split=2,
            min_weight_fraction_leaf=0.0, n_estimators=100, n_jobs=1,
            oob_score=False, random_state=None, verbose=0,
            warm_start=False)
```

In [465]:
```
forest.score(train_x_tfidf_array,train_y)
```

Out[465]:
```
0.92149999999999999
```

In [466]:
```
forest.score(test_x_tfidf_array,test_y)
```

Out[466]:
```
0.78949999999999998
```

In [312]:
```
forest_predictions = forest.predict(test_x_tfidf_array)
```

In [467]:
```
from sklearn.metrics import confusion_matrix
```

In [468]:
```
confusion_matrix(test_y, forest_predictions)
```

Out[468]:
```
array([[766, 234],
       [181, 819]])
```

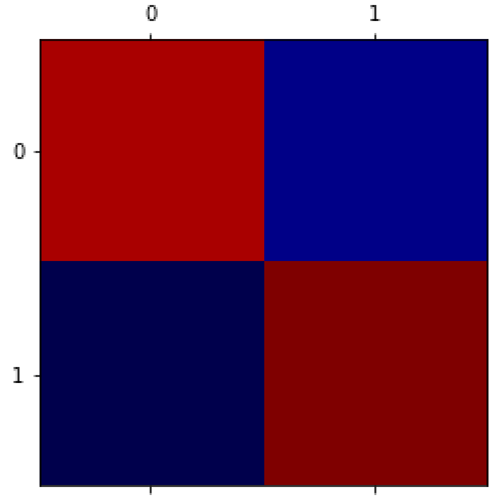
In [315]:
```
(766+819)/(766+819+234+181)
```

Out[315]:
```
0.7925
```

In [469]:
```
import matplotlib.pyplot as plt
```

In [470]:
```
plt.matshow(confusion_matrix(test_y, forest_predictions),cmap='seismic')
```

Out[470]:
```
<matplotlib.image.AxesImage at 0x13412f470>
```



In [ ]: