510 MBTI dataset- 1000 words RFECV

April 30, 2018

In [70]:
```
#import libraries
import numpy as np
import pandas as pd
import sklearn

from sklearn import preprocessing

from sklearn.feature_extraction.text import CountVectorizer

import matplotlib.pyplot as plt
plt.rc("font", size=14)
from sklearn.linear_model import LogisticRegression
from sklearn.cross_validation import train_test_split
import seaborn as sns
sns.set(style="white")
sns.set(style="whitegrid", color_codes=True)

import statsmodels.api as sm
```

In [71]:
```
#import dataset
data= pd.read_csv(r'C:\Users\Robert\Desktop\mbti_1.csv')
```

In [72]:
```
data.head()
```

Out[72]:

	type	posts
0	INFJ	'http://www.youtube.com/watch?v=qsXHcwe3krw|||...
1	ENTP	'I'm finding the lack of me in these posts ver...
2	INTP	'Good one _____ https://www.youtube.com/wat...
3	INTJ	'Dear INTP, I enjoyed our conversation the o...
4	ENTJ	'You're fired.|||That's another silly misconce...

In [73]:
```
#Add a column to specify if extroverted.
data['extroverted']= data.type.str[:1]
#df['New_Sample'] = df.Sample.str[:1]

data.head()
```

Out[73]:

	type	posts	extroverted
0	INFJ	'http://www.youtube.com/watch?v=qsXHcwe3krw|||...	I

```
1  ENTP  'I'm finding the lack of me in these posts ver...          E
2  INTP  'Good one  _____   https://www.youtube.com/wat...          I
3  INTJ  'Dear INTP,   I enjoyed our conversation the o...          I
4  ENTJ  'You're fired.|||That's another silly misconce...          E
```

In [74]:
```
#how many is I vs. E in dataset?
data.describe()
```

Out[74]:

	type	posts	extroverted
count	8675	8675	8675
unique	16	8675	2
top	INFP	'Afterburner your reasoning is EPIC!|||Patriot...	I
freq	1832	1	6676

In [75]:
```
#breakdown of types in survery
type_data = data.groupby('type')
type_data.describe()
```

Out[75]:

	extroverted				posts			top	freq
	count	unique	top	freq	count	unique			
type									
ENFJ	190	1	E	190	190	190		'There are some really awesome ENFJ facebook g...	1
ENFP	675	1	E	675	675	675		'What do you mean? What's changed? I feel a ...	1
ENTJ	231	1	E	231	231	231		'Usually when I am in a group consisting large...	1
ENTP	685	1	E	685	685	685		Haven't had time to think.|||Oh, christ. Now I...	1
ESFJ	42	1	E	42	42	42		Entj|||Esfp|||Entp|||Esfj|||Estp|||Infp|||Intj...	1
ESFP	48	1	E	48	48	48		'Good job! William I am!!!|||Yes to both. Sel...	1
ESTJ	39	1	E	39	39	39		hitler was what he was,and i am estj or esfj. ...	1
ESTP	89	1	E	89	89	89		'Class clown. I made a joke out of everything|...	1
INFJ	1470	1	I	1470	1470	1470			
INFP	1832	1	I	1832	1832	1832			
INTJ	1091	1	I	1091	1091	1091			
INTP	1304	1	I	1304	1304	1304			
ISFJ	166	1	I	166	166	166			
ISFP	271	1	I	271	271	271			
ISTJ	205	1	I	205	205	205			
ISTP	337	1	I	337	337	337			

```
INFJ  'desperately wish there was a moment every day...    1
INFP  'I can say if I was winked at I would be throw...    1
INTJ  'Afterburner your reasoning is EPIC!|||Patriot...    1
INTP  'I totally analyzed you a figured out who you ...    1
ISFJ  I would say Gayle is an INFP 4w5 so/sx, a very...    1
ISFP  'I was with an ENFJ, and this is extremely acc...    1
ISTJ  'http://www.youtube.com/watch?v=EOgfZHxTgts|||...    1
ISTP  'Lol. But hey, men can be gorgeous too.|||Nope...    1
```

In [76]:
```
data['extroverted'].replace(['I','E'],['0','1'],inplace=True)

data.head()
```

Out[76]:

	type	posts	extroverted
0	INFJ	'http://www.youtube.com/watch?v=qsXHcwe3krw|||...	0
1	ENTP	'I'm finding the lack of me in these posts ver...	1
2	INTP	'Good one _____ https://www.youtube.com/wat...	0
3	INTJ	'Dear INTP, I enjoyed our conversation the o...	0
4	ENTJ	'You're fired.|||That's another silly misconce...	1

In [77]:
```
data.dtypes
```

Out[77]:
```
type           object
posts          object
extroverted    object
dtype: object
```

In [78]:
```
data['extroverted'] = data['extroverted'].astype('int')
```

In [79]:
```
data.dtypes
```

Out[79]:
```
type           object
posts          object
extroverted     int32
dtype: object
```

In [80]:
```
from sklearn.feature_extraction import text
stop_words = text.ENGLISH_STOP_WORDS.union(['http', 'isfj', 'infp', 'intj', 'https',
                                            'infj', 'infp', 'intj','intp',
 'intp','istj', 'istp', '00', 'enfps', 'entps', 'infjs', 'enfjs', 'estps',
                                            'entj', 'esfjs', 'existence', 'infps', 'en

 '000',
 '01',
 '02',
 '03',
 '04',
 '05',
 '06',
 '07',
```

```
'08',
'09',
'10',
'100',
'1000',
'101',
'10th',
'11',
'110',
'11th',
'12',
'120',
'13',
'130',
'14',
'140',
'15',
'150',
'16',
'16personalities',
'17',
'18',
'180',
'19',
'1984',
'1995',
'1s',
'1st',
'1w2',
'1w9',
'20',
'200',
'2000',
'2001',
'2005',
'2006',
'2007',
'2008',
'2009',
'2010',
'2011',
'2012',
'2013',
'2014',
'2015',
'2016',
'2017',
'20s',
```

```
'20th',
'21',
'21st',
'22',
'23',
'24',
'25',
'26',
'27',
'28',
'29',
'2nd',
'2w1',
'2w3',
'30',
'300',
'30s',
'31',
'32',
'33',
'34',
'35',
'36',
'37',
'38',
'39',
'3D',
'3rd',
'3s',
'3w2',
'3w4',
'40',
'400',
'41',
'42',
'43',
'44',
'45',
'46',
'47',
'48',
'49',
'4s',
'4th',
'4w3',
'4w5',
'50',
'500',
```

```
'51',
'52',
'53',
'54',
'55',
'56',
'564x',
'57',
'58',
'59',
'5s',
'5th',
'5w4',
'5w6',
'60',
'600',
'60s',
'61',
'62',
'63',
'64',
'65',
'66',
'67',
'68',
'69',
'6s',
'6th',
'6w5',
'6w7',
'70',
'70s',
'71',
'72',
'73',
'736x',
'74',
'75',
'76',
'77',
'78',
'79',
'7s',
'7th',
'7w6',
'7w8',
'80',
'80s',
```

```
            '81',
            '82',
            '83',
            '84',
            '85',
            '86',
            '87',
            '88',
            '89',
            '8s',
            '8th',
            '8w7',
            '8w9',
            '90',
            '90s',
            '91',
            '92',
            '94',
            '95',
            '98',
            '99',
            '9s',
            '9th',
            '9w1',
            '9w8', '125',
            '160',
            '2003',
            '2004',])
```

In [93]:
```
#create our count vectorizer
count_vectorizer = CountVectorizer(analyzer='word', min_df=1, stop_words = stop_words

#Transform the list of strings
transform = count_vectorizer.fit_transform(data.posts).toarray()
```

In [94]:
```
#shows list of stopwords
#count_vectorizer.get_stop_words()
```

Out[94]:
```
frozenset({'00',
           '000',
           '01',
           '02',
           '03',
           '04',
           '05',
           '06',
           '07',
           '08',
```

```
'09',
'10',
'100',
'1000',
'101',
'10th',
'11',
'110',
'11th',
'12',
'120',
'125',
'13',
'130',
'14',
'140',
'15',
'150',
'16',
'160',
'16personalities',
'17',
'18',
'180',
'19',
'1984',
'1995',
'1s',
'1st',
'1w2',
'1w9',
'20',
'200',
'2000',
'2001',
'2003',
'2004',
'2005',
'2006',
'2007',
'2008',
'2009',
'2010',
'2011',
'2012',
'2013',
'2014',
'2015',
```

```
'2016',
'2017',
'20s',
'20th',
'21',
'21st',
'22',
'23',
'24',
'25',
'26',
'27',
'28',
'29',
'2nd',
'2w1',
'2w3',
'30',
'300',
'30s',
'31',
'32',
'33',
'34',
'35',
'36',
'37',
'38',
'39',
'3D',
'3rd',
'3s',
'3w2',
'3w4',
'40',
'400',
'41',
'42',
'43',
'44',
'45',
'46',
'47',
'48',
'49',
'4s',
'4th',
'4w3',
```

```
'4w5',
'50',
'500',
'51',
'52',
'53',
'54',
'55',
'56',
'564x',
'57',
'58',
'59',
'5s',
'5th',
'5w4',
'5w6',
'60',
'600',
'60s',
'61',
'62',
'63',
'64',
'65',
'66',
'67',
'68',
'69',
'6s',
'6th',
'6w5',
'6w7',
'70',
'70s',
'71',
'72',
'73',
'736x',
'74',
'75',
'76',
'77',
'78',
'79',
'7s',
'7th',
'7w6',
```

```
'7w8',
'80',
'80s',
'81',
'82',
'83',
'84',
'85',
'86',
'87',
'88',
'89',
'8s',
'8th',
'8w7',
'8w9',
'90',
'90s',
'91',
'92',
'94',
'95',
'98',
'99',
'9s',
'9th',
'9w1',
'9w8',
'a',
'about',
'above',
'across',
'after',
'afterwards',
'again',
'against',
'all',
'almost',
'alone',
'along',
'already',
'also',
'although',
'always',
'am',
'among',
'amongst',
'amoungst',
```

```
'amount',
'an',
'and',
'another',
'any',
'anyhow',
'anyone',
'anything',
'anyway',
'anywhere',
'are',
'around',
'as',
'at',
'back',
'be',
'became',
'because',
'become',
'becomes',
'becoming',
'been',
'before',
'beforehand',
'behind',
'being',
'below',
'beside',
'besides',
'between',
'beyond',
'bill',
'both',
'bottom',
'but',
'by',
'call',
'can',
'cannot',
'cant',
'co',
'com',
'con',
'could',
'couldnt',
'cry',
'de',
'describe',
```

```
'detail',
'do',
'done',
'down',
'due',
'during',
'each',
'eg',
'eight',
'either',
'eleven',
'else',
'elsewhere',
'empty',
'enfj',
'enfjs',
'enfp',
'enfps',
'enough',
'entj',
'entjs',
'entp',
'entps',
'esfjs',
'estps',
'etc',
'even',
'ever',
'every',
'everyone',
'everything',
'everywhere',
'except',
'existence',
'few',
'fifteen',
'fifty',
'fill',
'find',
'fire',
'first',
'five',
'for',
'former',
'formerly',
'forty',
'found',
'four',
```

```
'from',
'front',
'full',
'further',
'get',
'give',
'go',
'had',
'has',
'hasnt',
'have',
'he',
'hence',
'her',
'here',
'hereafter',
'hereby',
'herein',
'hereupon',
'hers',
'herself',
'him',
'himself',
'his',
'how',
'however',
'http',
'https',
'hundred',
'i',
'ie',
'if',
'in',
'inc',
'indeed',
'infj',
'infjs',
'infp',
'infps',
'interest',
'intj',
'into',
'intp',
'intps',
'is',
'isfj',
'istj',
'istp',
```

```
'it',
'its',
'itself',
'keep',
'last',
'latter',
'latterly',
'least',
'less',
'ltd',
'made',
'many',
'may',
'me',
'meanwhile',
'might',
'mill',
'mine',
'more',
'moreover',
'most',
'mostly',
'move',
'much',
'must',
'my',
'myself',
'name',
'namely',
'neither',
'never',
'nevertheless',
'next',
'nine',
'no',
'nobody',
'none',
'noone',
'nor',
'not',
'nothing',
'now',
'nowhere',
'of',
'off',
'often',
'on',
'once',
```

```
'one',
'only',
'onto',
'or',
'other',
'others',
'otherwise',
'our',
'ours',
'ourselves',
'out',
'over',
'own',
'part',
'per',
'perhaps',
'please',
'put',
'rather',
're',
'same',
'see',
'seem',
'seemed',
'seeming',
'seems',
'serious',
'several',
'she',
'should',
'show',
'side',
'since',
'sincere',
'six',
'sixty',
'so',
'some',
'somehow',
'someone',
'something',
'sometime',
'sometimes',
'somewhere',
'still',
'such',
'system',
'take',
```

```
'ten',
'than',
'that',
'the',
'their',
'them',
'themselves',
'then',
'thence',
'there',
'thereafter',
'thereby',
'therefore',
'therein',
'thereupon',
'these',
'they',
'thick',
'thin',
'third',
'this',
'those',
'though',
'three',
'through',
'throughout',
'thru',
'thus',
'to',
'together',
'too',
'top',
'toward',
'towards',
'twelve',
'twenty',
'two',
'un',
'under',
'until',
'up',
'upon',
'us',
'very',
'via',
'was',
'we',
'well',
```

```
                    'were',
                    'what',
                    'whatever',
                    'when',
                    'whence',
                    'whenever',
                    'where',
                    'whereafter',
                    'whereas',
                    'whereby',
                    'wherein',
                    'whereupon',
                    'wherever',
                    'whether',
                    'which',
                    'while',
                    'whither',
                    'who',
                    'whoever',
                    'whole',
                    'whom',
                    'whose',
                    'why',
                    'will',
                    'with',
                    'within',
                    'without',
                    'would',
                    'yet',
                    'you',
                    'your',
                    'yours',
                    'yourself',
                    'yourselves',
                    'youtube'})
```

```
In [95]: transform
```

```
Out[95]: array([[0, 0, 0, ..., 0, 0, 0],
                [0, 1, 0, ..., 0, 0, 0],
                [2, 1, 2, ..., 0, 0, 0],
                ...,
                [0, 0, 0, ..., 0, 0, 1],
                [0, 2, 0, ..., 0, 0, 0],
                [0, 1, 0, ..., 1, 0, 0]], dtype=int64)
```

```
In [96]: transform.shape
```

```
Out[96]: (8675, 1000)
```

```
In [97]: count_vectorizer.get_feature_names()
```

```
Out[97]: ['ability',
          'able',
          'absolutely',
          'accept',
          'according',
          'accurate',
          'act',
          'action',
          'actions',
          'actual',
          'actually',
          'add',
          'admit',
          'advice',
          'afraid',
          'age',
          'ago',
          'agree',
          'ah',
          'albums',
          'alot',
          'alright',
          'amazing',
          'anger',
          'angry',
          'animals',
          'anime',
          'annoying',
          'answer',
          'answers',
          'anxiety',
          'anymore',
          'anyways',
          'apparently',
          'appreciate',
          'approach',
          'area',
          'aren',
          'argument',
          'art',
          'article',
          'ask',
          'asked',
          'asking',
          'ass',
          'assume',
```

```
'attention',
'attracted',
'attractive',
'avatar',
'average',
'avoid',
'aware',
'away',
'awesome',
'awkward',
'baby',
'bad',
'based',
'basically',
'beautiful',
'bed',
'beginning',
'behavior',
'believe',
'best',
'better',
'big',
'biggest',
'bit',
'black',
'blue',
'blushed',
'body',
'book',
'books',
'bored',
'boring',
'born',
'bother',
'boy',
'boyfriend',
'brain',
'break',
'bring',
'brother',
'btw',
'bunch',
'business',
'busy',
'buy',
'called',
'calm',
'came',
```

```
'car',
'care',
'career',
'caring',
'case',
'cat',
'cats',
'cause',
'certain',
'certainly',
'chance',
'change',
'changed',
'character',
'characters',
'check',
'child',
'childhood',
'children',
'choice',
'choose',
'city',
'class',
'classes',
'clear',
'clearly',
'close',
'closest',
'coffee',
'cognitive',
'cold',
'college',
'color',
'come',
'comes',
'comfortable',
'coming',
'comment',
'comments',
'common',
'communication',
'company',
'complete',
'completely',
'computer',
'concept',
'confidence',
'confident',
```

```
'conflict',
'confused',
'connection',
'consider',
'considered',
'considering',
'constantly',
'contact',
'content',
'control',
'conversation',
'conversations',
'cool',
'correct',
'couldn',
'count',
'country',
'couple',
'course',
'crap',
'crazy',
'create',
'creative',
'crying',
'curious',
'current',
'currently',
'cut',
'cute',
'dad',
'damn',
'dark',
'date',
'dated',
'dating',
'day',
'days',
'dead',
'deal',
'dealing',
'dear',
'death',
'debate',
'decide',
'decided',
'decision',
'decisions',
'deep',
```

```
'definitely',
'definition',
'degree',
'depending',
'depends',
'depressed',
'depression',
'described',
'description',
'descriptions',
'desire',
'despite',
'details',
'developed',
'did',
'didn',
'die',
'difference',
'different',
'difficult',
'disagree',
'discussion',
'dislike',
'does',
'doesn',
'dog',
'doing',
'dom',
'dominant',
'don',
'dont',
'door',
'double',
'doubt',
'draw',
'dream',
'dreams',
'drink',
'drive',
'drunk',
'dry',
'dude',
'earlier',
'early',
'earth',
'easier',
'easily',
'easy',
```

```
'eat',
'eating',
'edit',
'effort',
'emotion',
'emotional',
'emotionally',
'emotions',
'end',
'ended',
'energy',
'english',
'enjoy',
'enneagram',
'entire',
'entirely',
'environment',
'esfj',
'esfp',
'especially',
'estj',
'estp',
'eventually',
'everybody',
'evil',
'ex',
'exact',
'exactly',
'example',
'excited',
'exist',
'expect',
'experience',
'experienced',
'experiences',
'explain',
'express',
'extremely',
'extrovert',
'extroverted',
'eye',
'eyes',
'face',
'facebook',
'fact',
'fair',
'fairly',
'fall',
```

```
'family',
'fan',
'far',
'fast',
'father',
'favorite',
'favourite',
'fe',
'fear',
'feature',
'feel',
'feeling',
'feelings',
'feels',
'fellow',
'felt',
'female',
'fi',
'fight',
'figure',
'figured',
'finally',
'finding',
'fine',
'fit',
'fits',
'fix',
'focus',
'follow',
'following',
'food',
'forever',
'forget',
'forgot',
'form',
'forum',
'forums',
'forward',
'free',
'friend',
'friendly',
'friends',
'friendship',
'frustrating',
'fuck',
'fucking',
'fully',
'fun',
```

```
'function',
'functions',
'funny',
'future',
'game',
'games',
'gave',
'gender',
'general',
'generally',
'gets',
'getting',
'gif',
'girl',
'girlfriend',
'girls',
'given',
'gives',
'giving',
'glad',
'goal',
'god',
'goes',
'going',
'gone',
'gonna',
'good',
'got',
'gotten',
'great',
'green',
'group',
'grow',
'growing',
'guess',
'guy',
'guys',
'ha',
'haha',
'hahaha',
'hair',
'half',
'hand',
'handle',
'hands',
'hang',
'happen',
'happened',
```

```
'happens',
'happy',
'hard',
'hate',
'haven',
'having',
'head',
'healthy',
'hear',
'heard',
'heart',
'hell',
'hello',
'help',
'helped',
'helpful',
'helps',
'hey',
'hi',
'high',
'higher',
'highly',
'history',
'hit',
'hmm',
'hold',
'home',
'honest',
'honestly',
'hope',
'horrible',
'hot',
'hour',
'hours',
'house',
'hug',
'huge',
'hugs',
'human',
'humans',
'humor',
'hurt',
'husband',
'idea',
'ideal',
'ideas',
'identify',
'ignore',
```

```
'im',
'image',
'images',
'imagine',
'imgur',
'important',
'impossible',
'incredibly',
'individual',
'inferior',
'information',
'inner',
'inside',
'instead',
'intelligence',
'intelligent',
'intense',
'interested',
'interesting',
'interests',
'internet',
'intjs',
'introvert',
'introverted',
'introverts',
'intuition',
'intuitive',
'involved',
'iphone',
'isfp',
'isn',
'issue',
'issues',
'istps',
'ive',
'job',
'joke',
'jpg',
'judging',
'just',
'kid',
'kids',
'kill',
'kind',
'kinda',
'kitteh',
'knew',
'know',
```

```
'knowing',
'knowledge',
'known',
'knows',
'lack',
'language',
'large',
'late',
'lately',
'later',
'laugh',
'laughing',
'lazy',
'lead',
'learn',
'learned',
'learning',
'leave',
'left',
'let',
'level',
'lie',
'life',
'light',
'like',
'liked',
'likely',
'likes',
'line',
'link',
'list',
'listen',
'listening',
'literally',
'little',
'live',
'lived',
'lives',
'living',
'll',
'logic',
'logical',
'lol',
'lonely',
'long',
'longer',
'look',
'looked',
```

```
'looking',
'looks',
'lose',
'lost',
'lot',
'lots',
'loud',
'love',
'loved',
'loves',
'low',
'mad',
'main',
'mainly',
'major',
'majority',
'make',
'makes',
'making',
'male',
'man',
'married',
'match',
'math',
'matter',
'maybe',
'mbti',
'mean',
'meaning',
'means',
'meant',
'media',
'meet',
'meeting',
'men',
'mental',
'mention',
'mentioned',
'message',
'met',
'middle',
'mind',
'minutes',
'miss',
'mom',
'moment',
'money',
'month',
```

```
'months',
'mood',
'morning',
'mother',
'moved',
'movie',
'movies',
'moving',
'music',
'natural',
'naturally',
'nature',
'ne',
'near',
'nearly',
'necessarily',
'need',
'needed',
'needs',
'negative',
'net',
'new',
'nf',
'ni',
'nice',
'night',
'non',
'nope',
'normal',
'normally',
'note',
'notice',
'noticed',
'nt',
'number',
'obvious',
'obviously',
'odd',
'oh',
'ok',
'okay',
'old',
'older',
'omg',
'ones',
'online',
'op',
'open',
```

```
'opinion',
'opinions',
'opposite',
'order',
'original',
'outside',
'page',
'pain',
'parents',
'particular',
'particularly',
'partner',
'parts',
'party',
'past',
'pay',
'people',
'perc',
'perfect',
'perfectly',
'person',
'personal',
'personality',
'personalitycafe',
'personally',
'perspective',
'philosophy',
'phone',
'physical',
'pick',
'picture',
'pictures',
'place',
'places',
'plan',
'play',
'played',
'playing',
'plus',
'png',
'point',
'points',
'poor',
'position',
'positive',
'possible',
'possibly',
'post',
```

```
'posted',
'posting',
'posts',
'potential',
'power',
'prefer',
'present',
'pretty',
'probably',
'problem',
'problems',
'process',
'profile',
'project',
'proud',
'psychology',
'public',
'purpose',
'putting',
'quality',
'question',
'questions',
'quick',
'quickly',
'quiet',
'quite',
'quote',
'random',
'rare',
'rarely',
'read',
'reading',
'ready',
'real',
'reality',
'realize',
'realized',
'really',
'reason',
'reasons',
'recently',
'red',
'relate',
'related',
'relationship',
'relationships',
'religion',
'religious',
```

```
'remember',
'reminds',
'reply',
'research',
'respect',
'respond',
'response',
'responses',
'rest',
'result',
'results',
'right',
'rock',
'romantic',
'room',
'rules',
'run',
'running',
'sad',
'said',
'saw',
'say',
'saying',
'says',
'scared',
'school',
'science',
'score',
'se',
'second',
'seeing',
'seen',
'self',
'sense',
'sensitive',
'sent',
'series',
'seriously',
'set',
'sex',
'sexual',
'share',
'sharing',
'shit',
'shocked',
'short',
'shouldn',
'shows',
```

```
'shy',
'si',
'sick',
'sign',
'silly',
'similar',
'simple',
'simply',
'single',
'sister',
'sit',
'site',
'sitting',
'situation',
'situations',
'skills',
'sleep',
'slightly',
'small',
'smart',
'smile',
'social',
'socially',
'society',
'socionics',
'somebody',
'somewhat',
'song',
'songs',
'soon',
'sorry',
'sort',
'soul',
'sound',
'sounds',
'sp',
'space',
'speak',
'speaking',
'special',
'specific',
'specifically',
'spend',
'spent',
'spot',
'stand',
'start',
'started',
```

```
'starting',
'state',
'statement',
'stay',
'step',
'stick',
'stop',
'stories',
'story',
'straight',
'strange',
'stress',
'strong',
'stuck',
'study',
'studying',
'stuff',
'stupid',
'style',
'subject',
'sucks',
'suggest',
'summer',
'super',
'support',
'suppose',
'supposed',
'sure',
'surprised',
'sweet',
'sx',
'taken',
'takes',
'taking',
'talk',
'talked',
'talking',
'tapatalk',
'te',
'teacher',
'tell',
'telling',
'tend',
'term',
'terms',
'terrible',
'test',
'tested',
```

```
'tests',
'text',
'thank',
'thanks',
'thats',
'theory',
'thing',
'things',
'think',
'thinking',
'thinks',
'thought',
'thoughts',
'thread',
'threads',
'ti',
'time',
'times',
'tired',
'title',
'today',
'told',
'tongue',
'took',
'topic',
'totally',
'touch',
'traits',
'tried',
'trouble',
'true',
'truly',
'trust',
'truth',
'try',
'trying',
'tumblr',
'turn',
'turned',
'tv',
'type',
'typed',
'types',
'typical',
'typically',
'typing',
'uncomfortable',
'understand',
```

```
'understanding',
'understood',
'unfortunately',
'unhealthy',
'universe',
'university',
'unless',
'unsure',
'uploads',
'upset',
'use',
'used',
'useful',
'user',
'using',
'usually',
'value',
'values',
've',
'version',
'video',
'videos',
'view',
'voice',
'vs',
'wait',
'waiting',
'walk',
'walking',
'wanna',
'want',
'wanted',
'wanting',
'wants',
'warm',
'wasn',
'watch',
'watched',
'watching',
'water',
'way',
'ways',
'wear',
'week',
'weeks',
'weird',
'welcome',
'went',
```

```
 'white',
 'wife',
 'willing',
 'win',
 'wink',
 'wish',
 'woman',
 'women',
 'won',
 'wonder',
 'wonderful',
 'wondering',
 'word',
 'words',
 'work',
 'worked',
 'working',
 'works',
 'world',
 'worry',
 'worse',
 'worst',
 'worth',
 'wouldn',
 'wow',
 'write',
 'writing',
 'written',
 'wrong',
 'wrote',
 'www',
 'xd',
 'ya',
 'yeah',
 'year',
 'years',
 'yep',
 'yes',
 'yesterday',
 'young',
 'younger',
 'youtu']
```

In [98]:
```
#deploy and evaluate model
X=transform
y= data.extroverted
LogReg = LogisticRegression()
log_model = LogReg.fit(X, y)
```

```
print(LogReg.score(X, y))
```

```
0.831469740634
```

```
In [99]: y_pred = LogReg.predict(X)
         from sklearn.metrics import classification_report
         #print(classification_report(y, y_pred))
         y_pred
```

```
Out[99]: array([0, 1, 0, ..., 0, 0, 0])
```

```
In [100]: #predict probability
          LogReg.predict_proba(X)
```

```
Out[100]: array([[ 0.98611172,  0.01388828],
                 [ 0.24652963,  0.75347037],
                 [ 0.9876008 ,  0.0123992 ],
                 ...,
                 [ 0.79772296,  0.20227704],
                 [ 0.58838976,  0.41161024],
                 [ 0.91377169,  0.08622831]])
```

```
In [101]: # Check trained model intercept
          print("Intercept is ",  log_model.intercept_)

          # Check trained model coefficients
          print("Coefficients are",  log_model.coef_)
```

```
Intercept is  [-0.87775154]
Coefficients are [[ -5.28997687e-02   2.86176034e-02   3.36041054e-02  -2.36611411e-02
    1.00055925e-02  -1.36921426e-01   2.17613577e-02  -1.52126704e-02
    2.30634796e-01   4.47117324e-02  -4.68193753e-05  -2.16254118e-03
    5.37954702e-02   1.87829004e-02  -2.01270984e-01  -8.77103720e-03
    1.06312899e-01  -1.03347004e-02  -4.69906221e-02   3.34363028e-02
    9.54529808e-02  -6.75883567e-02   8.91682435e-02  -4.90272948e-02
    3.02870909e-02  -1.30802291e-01  -3.02986439e-01   1.20644477e-01
    2.56006265e-02  -1.95581450e-01  -5.05059479e-02   1.87568946e-02
   -5.99876815e-02  -4.05848147e-02   1.25755964e-02   7.65851671e-02
    1.07088056e-01  -2.27790255e-02   9.31370152e-02  -3.21706362e-02
   -4.04155789e-02   1.43227058e-02  -1.61132109e-03  -1.54815822e-01
    6.71914048e-02   9.74678338e-03   1.57335691e-01  -4.94103679e-02
   -9.27600191e-02  -1.46619570e-02  -2.19002404e-01   4.54936003e-02
    7.77545383e-02  -4.77031209e-02   1.19555513e-01  -1.44268285e-01
    9.59875856e-02  -2.81392539e-02   3.40303975e-03   6.42559876e-02
   -1.10947110e-01   6.84217302e-02   1.71915499e-01   2.85682261e-02
    2.07437915e-02   2.03683336e-02  -2.32612797e-02   3.96799149e-02
    1.00810903e-01   4.05735170e-02   1.48017034e-02   1.34121369e-01
    3.56266200e-02  -3.84127276e-03   5.25155633e-02  -1.72263712e-01
```

```
 3.49278151e-01   2.24521159e-01  -1.23025910e-01  -9.17325055e-02
-2.25084428e-01   1.48371483e-02   4.07065339e-02   1.82871832e-02
 3.91179505e-02   4.25535807e-03   2.61004935e-02   4.28101831e-02
 3.05046564e-01   1.88881989e-01  -2.28283675e-02  -9.06765282e-02
-8.86773899e-02   1.52714797e-02  -2.76573060e-02  -1.37416022e-02
 1.06736168e-01  -1.06547841e-01   6.68324882e-02  -5.24002303e-02
-2.10862957e-01   9.71836411e-02   4.22211990e-03  -6.38561180e-02
 1.43793559e-02   1.34719716e-02   2.27437089e-03  -6.17419088e-02
-1.05678449e-01  -1.23223700e-01   4.32514485e-02  -7.44853128e-03
-1.06485214e-01   1.17896633e-01  -3.28993416e-03  -1.13651584e-01
-3.01435710e-02  -1.19039844e-01   1.07147329e-01   1.25448503e-04
-8.84103457e-02  -5.16399222e-02   9.17302427e-02  -3.75459538e-02
-1.18950910e-01  -1.73848263e-02  -5.19033280e-02  -4.39949961e-02
 1.88395830e-02  -2.32578701e-01  -1.00172558e-01  -2.22312107e-02
-1.48666646e-01   5.99865889e-02   1.18668193e-01  -4.32615817e-02
 1.12053530e-01   1.05178869e-01  -1.74916493e-01  -1.03589556e-01
 7.05573886e-02  -1.01184408e-01  -5.59910140e-02  -1.32428807e-01
 6.32091856e-02   8.66428913e-02   4.70441662e-02  -9.93923358e-03
 3.18818066e-02   6.20595706e-02  -1.27411201e-01   5.21032751e-02
-3.71788734e-02   7.86555719e-02   3.26012346e-03   1.53314270e-01
-7.78600768e-02   2.51807839e-02  -5.84667501e-03  -1.09733087e-01
-1.12785492e-02   6.86634706e-02   1.27652754e-01  -1.41873331e-01
-4.80624260e-02   1.18134678e-01   6.65448469e-02  -8.29321651e-02
 7.69192906e-02  -1.50119899e-01   1.49314465e-01  -4.26906307e-02
 7.01428625e-02  -2.33066531e-02   5.41609810e-02   2.65953946e-01
-4.47861307e-02  -3.98040186e-02   4.91250475e-02  -1.20844892e-01
 1.73698732e-02  -3.51535610e-01  -1.59698212e-02  -1.54738902e-01
 3.00927773e-01  -7.84649452e-02   5.96375770e-02   2.60979619e-01
 4.39491333e-02  -5.04632776e-02   6.17545654e-02  -2.52130022e-03
-5.04022983e-02  -1.27600315e-01   7.29031911e-02  -2.01890322e-02
 2.34038608e-03   1.19140829e-01  -3.20973827e-02  -6.15091232e-03
 1.62471427e-01  -2.00561426e-01  -2.62644446e-01   3.04342534e-01
 3.08036589e-05  -2.22960959e-02  -2.11846301e-02  -6.73040407e-02
-1.29697059e-02  -1.37635170e-01  -1.76393781e-01   9.95264509e-02
-2.03733271e-01   2.77558357e-03  -1.14404846e-02  -9.62933240e-02
-1.68036153e-02  -3.59238550e-02   6.75289722e-02  -2.84192248e-02
-1.06972614e-02  -8.46093007e-02  -3.53814924e-02  -1.95564437e-02
-1.06839988e-01  -2.24845220e-01   6.13106671e-03  -1.18584635e-01
 1.01687657e-01  -4.78851511e-02  -3.26158535e-01  -1.51405338e-02
-1.48267708e-01  -1.16816803e-01  -2.98882453e-01   1.62018296e-01
 8.38859604e-02   3.16902983e-02  -5.01432140e-02   5.54868128e-02
-4.17403741e-02   1.12354438e-01   1.60715350e-02  -7.07085284e-02
-1.58284491e-01  -7.01113363e-02   2.53192676e-02  -1.78838032e-01
 7.53072285e-02  -7.32896037e-02  -1.21431286e-02   4.58879265e-02
-1.28819852e-01   1.61197337e-02  -2.29718955e-02   4.89691901e-02
 8.86352671e-02   1.53595363e-02   7.73073186e-02   1.76229995e-01
-1.23871551e-03  -8.26497126e-02   3.99721611e-02  -1.42970506e-02
 2.61089392e-01   1.65267742e-02  -1.00437686e-02   3.34440548e-01
```

```
-3.40570686e-02   4.04260946e-02   1.43219506e-02   1.55768151e-01
-1.50831037e-01   9.89817755e-02   1.08428735e-01   3.78839164e-02
 3.15658102e-01  -3.78772635e-02  -1.16082366e-01  -5.55160930e-02
 5.92230035e-02  -4.77602905e-02   2.35722754e-02  -9.74531367e-02
-1.72213073e-02  -3.47425598e-02  -1.80518444e-01   1.72404889e-01
 4.07678185e-02   2.18803422e-02  -1.06351917e-01   8.79393816e-02
 9.29220322e-02   6.28814390e-02   6.35283211e-03  -1.27344795e-02
-5.15926117e-02  -1.33511982e-01  -1.51836143e-04   3.44209068e-02
 2.54300247e-02   4.58661190e-02  -3.73447253e-02  -5.99384681e-02
 1.76994527e-01  -9.69980766e-03  -2.96471194e-02  -3.00121061e-02
-5.36745987e-02  -1.43589728e-02   8.03402071e-02   5.07948758e-02
-2.94765779e-02   6.24451748e-02   9.63240565e-02   4.78082609e-02
-3.26826012e-02  -6.49653272e-02  -1.26002530e-01   6.05243823e-02
-1.59826564e-01  -1.50535369e-02   1.02410379e-02  -2.32512549e-02
 3.63344188e-02  -4.76509378e-02   2.53204681e-01   1.17931202e-01
-1.06647928e-01  -2.07214624e-01   5.64966425e-02   1.55686022e-01
-1.06630791e-01   1.84761322e-01   8.26571854e-03  -8.24673163e-03
 7.57874502e-03   1.78104183e-01  -4.89445747e-02  -2.63471438e-01
 4.13791651e-02  -2.60612929e-01  -3.43025943e-02   4.99595528e-02
 9.11759561e-02   1.06408985e-02   2.74758538e-02  -3.39605974e-02
-1.14038960e-01  -9.46525694e-02  -3.73764986e-02   7.33513235e-02
 1.41532145e-01   8.73910571e-02   1.64947972e-01  -4.82205036e-03
 6.94268957e-02   1.49649241e-02   3.87727748e-02   3.41416534e-02
 6.17669398e-03   4.36344623e-02   3.62170944e-03   5.73629690e-02
-1.76801870e-01   1.91659565e-02   1.04193530e-01  -2.22561693e-05
-2.97182722e-02   5.65288697e-02   7.34550591e-02   7.84555242e-02
 5.91696000e-02   1.53884280e-01  -1.91484076e-02   8.21742777e-02
 3.07160547e-02   1.03853766e-01   1.38104945e-01   3.84421495e-03
 3.36489110e-02  -1.15587633e-01  -1.28235279e-02   2.71639173e-03
-5.40882145e-02  -1.03418304e-01  -6.35629056e-02   2.95004806e-02
-6.66122835e-02   3.13521192e-02  -7.00701933e-02  -1.16133218e-01
-4.26054210e-02   1.04109141e-01  -3.36390200e-02  -2.09379384e-03
 4.94742800e-02  -1.30342955e-02  -1.81165688e-02   5.40325301e-02
-4.90935275e-02  -2.06294037e-02   1.11930593e-01  -6.65822957e-02
 1.04464607e-01   3.69822107e-02  -1.61585075e-03   3.01334681e-02
-1.17807449e-01  -4.86984686e-02  -1.19977703e-02  -6.25882822e-02
-8.72232983e-02   1.83528291e-01  -2.18652586e-01   1.47518818e-01
-4.34099918e-02   4.83061899e-02   9.09503234e-02   6.74754266e-02
-9.15670606e-02   2.21189945e-02   3.26216337e-02   3.68361399e-02
 5.32245000e-02  -3.17810734e-02   3.36326693e-03   1.64297013e-01
-5.39729542e-02   1.04324060e-01  -2.95437526e-02  -8.87672460e-03
 4.01758732e-02  -1.27287887e-01  -1.03379346e-01  -1.49025350e-02
-6.91121340e-02   3.55796874e-02   3.15042191e-02  -2.86345256e-01
 3.17893005e-02  -6.11545189e-02  -1.20826887e-02  -1.61940681e-02
 1.21074630e-01  -7.79719105e-02   6.95344239e-02  -6.11279968e-03
 6.45249487e-02  -1.67525626e-02  -1.65311627e-01  -1.21299640e-01
-1.32443714e-01  -9.51262500e-03  -6.66672776e-02   1.74105195e-01
 7.27043675e-02   5.50605278e-02   1.42397860e-02  -1.24392991e-01
```

```
 4.10193734e-02  -1.92435570e-03   6.10674343e-02  -9.10525070e-02
-8.44513455e-02  -5.96593286e-02   4.22946907e-02  -3.91376724e-02
-1.81618015e-01  -8.39978736e-03  -2.44834387e-02   1.58894406e-02
-1.08004283e-03  -5.50080540e-02   5.82458148e-02   4.06688104e-02
-9.03479712e-02  -1.56151110e-02   4.30672076e-02   2.55520622e-02
 1.24270145e-01   3.44759683e-02  -6.85284125e-02  -9.39562082e-02
 1.81513017e-03   8.98382698e-02  -4.73581924e-02  -2.63689946e-02
 9.44420307e-02   3.04435648e-02  -3.41734249e-02   2.65628013e-01
-9.57070394e-02  -4.91201540e-02   5.73261591e-02   2.78896005e-02
-4.54053365e-02   1.07477156e-01   1.92637653e-02   1.55891156e-01
-6.25368473e-04  -1.02900667e-01   2.61388009e-02  -1.58005039e-01
 8.33250893e-02  -1.31921245e-01   1.36412408e-01  -1.97495552e-01
-1.70432995e-02   1.01614216e-02  -1.62507890e-01  -2.35575815e-02
-1.94965251e-02  -3.50431281e-02  -9.15997786e-02   4.58551594e-02
-8.68703848e-02   1.77707932e-02  -6.21700056e-02  -5.90695399e-02
 3.45992762e-02   9.87900404e-02   1.48895187e-02  -1.61006546e-01
-5.82180626e-02  -4.99537278e-02  -5.42484417e-02  -1.10583648e-01
 1.01078818e-01  -7.50802880e-02   1.54879585e-02   7.81367082e-02
 2.32053208e-03   5.12016827e-02  -5.65178031e-02   2.19573801e-01
-4.24718770e-02   2.36105657e-01  -1.33243792e-02  -1.35492357e-01
-7.11842261e-02   1.18963607e-01   2.03309414e-02  -4.11445071e-02
 3.24692736e-02  -4.07301664e-02  -2.60027747e-02   1.89898902e-03
 1.15913938e-01  -3.08554118e-02   6.60620309e-02  -5.48502325e-02
 6.73090944e-02   1.95898337e-02  -4.70445015e-02   1.27277658e-01
 2.88990318e-02   5.62886220e-02   5.81742354e-02   8.37502557e-02
-1.09313925e-01   1.59455634e-02  -5.27855171e-03  -1.57138826e-01
-6.62986181e-02   5.91642585e-02  -5.40164922e-02  -1.32750477e-01
 2.49757518e-02   1.31472403e-01   9.13845320e-02   5.88296697e-02
 1.02623937e-01   3.82131050e-02   9.06257892e-02   5.79640322e-02
-4.66548074e-02  -8.95003689e-02   6.67231542e-02   4.04829228e-02
 6.45360688e-02  -1.72131902e-02  -5.77477751e-02  -3.00300875e-02
 1.22916310e-01  -2.76107916e-01   3.37573073e-01   1.03837425e-02
-4.58709041e-02  -9.21122504e-02  -8.49942936e-03   2.87942795e-02
 5.75469223e-02  -6.83071263e-02  -1.60610671e-02   6.71843499e-02
-5.78803971e-02  -2.01411811e-01  -2.75328769e-02   2.90751186e-02
 2.67329488e-02  -5.38285882e-02   2.40232767e-02   1.13806716e-01
-1.79671742e-01  -7.60417275e-02   5.25447248e-03   2.08255012e-01
-2.68688502e-02  -1.26425903e-02   4.75788986e-02   2.38305534e-02
 7.46690857e-02   1.62394372e-02  -4.13293243e-02  -8.76628500e-02
 1.49864278e-03   1.16326556e-01   5.48603463e-02  -1.00878208e-01
-7.63121624e-02  -2.21252021e-02  -1.71870753e-02   9.62578086e-02
-5.40728631e-02   2.69756776e-02   1.40697086e-01  -6.86745384e-02
 1.37487039e-01  -4.71982742e-02   2.41348111e-02  -1.71936724e-01
-1.77083263e-01  -2.89594729e-02  -1.25903337e-01   3.54974348e-02
-5.19029422e-02  -1.79316955e-01   2.56333547e-02  -4.64878226e-02
-1.12795778e-01  -1.14499549e-01   4.62786878e-03   5.80063083e-02
 3.27264283e-02   1.00734318e-03   7.34799972e-02  -7.56822961e-02
-2.21045098e-01  -9.45230658e-02  -1.66506976e-02   5.86605604e-02
```

```
 7.92966008e-03  -4.23155309e-02   1.77597956e-02   1.49921359e-01
-3.34848868e-02   4.33954411e-02  -1.72693360e-01   2.89170552e-02
 4.24825074e-02  -5.07802059e-02   6.68433093e-02  -6.46543750e-02
 1.36751808e-01   3.76031927e-02   6.77013751e-02   2.34696498e-02
-1.30298725e-01  -1.54087547e-03  -5.62597950e-03  -3.53006823e-02
 6.16208996e-02   5.56443914e-02   3.28326450e-02   6.10645549e-04
-3.65160713e-02  -1.57103294e-02  -4.70208832e-02  -2.85469233e-02
-3.76163325e-02  -1.33178440e-01   8.12981408e-02   8.66328166e-02
 2.91392534e-02   7.81367226e-02  -1.64801012e-01  -5.33135238e-02
 9.87383456e-02  -4.26938784e-02   5.28458073e-02   1.42835076e-01
 3.76366544e-02   1.22231107e-01  -3.40919553e-01  -2.25771590e-02
-3.47094270e-02  -1.16453259e-01  -1.27368653e-01  -1.12757770e-01
 6.38961116e-02  -6.75105145e-03   1.81608813e-02  -4.06600199e-02
-1.05485990e-01  -5.57423599e-02   2.71169953e-02  -2.26349815e-03
-1.70121432e-02  -4.07651003e-02  -5.00098949e-02   2.44767465e-02
-7.37644875e-02   2.27810490e-03   2.78992810e-02   2.33698576e-02
-1.86302020e-01  -5.55820279e-02   5.32684919e-03  -6.73658503e-02
-2.56676504e-03  -3.60243559e-02   1.72487022e-02   2.86277289e-02
-3.65252276e-02   5.23365818e-02  -7.37428954e-02  -6.96555580e-02
-5.12401838e-02  -1.92671477e-02   1.01063244e-02  -9.54785894e-03
 1.97994112e-02  -3.58097641e-02   1.03273785e-01  -1.11895495e-01
-5.72141828e-02   8.81396118e-03  -1.53990755e-02  -1.97418997e-02
-2.12134911e-02  -9.14534940e-02  -9.29848048e-02  -2.01688108e-02
-1.51939596e-01   1.03931477e-01  -1.53199485e-01  -2.24810471e-02
-7.87769685e-02   3.41436255e-02  -4.52010559e-02   3.86462005e-02
-1.18917025e-01   2.64468040e-01   1.95523715e-02   2.29814848e-02
 1.46755783e-02   4.49838613e-02   7.18343673e-02   2.80078524e-02
 3.69140928e-02   2.06892973e-02   2.01144876e-03   6.89179562e-02
 6.94996052e-02   8.97421531e-03   1.33310308e-01  -9.99103754e-02
-4.15807730e-02  -3.29267516e-01  -1.71465586e-02  -9.06390900e-02
 5.23938274e-04  -1.45855650e-02  -9.35074496e-02  -8.71206246e-02
-1.45039094e-01  -2.97254071e-02  -1.28025459e-01   4.26457118e-02
-9.50446972e-02   9.73810008e-02  -4.00517421e-03  -5.52834124e-02
-1.70970106e-01  -9.38058291e-03  -1.27209498e-01  -2.46269675e-02
-8.70215499e-02  -4.59703066e-02   7.22280508e-02   8.47029856e-02
-7.78571969e-02   1.76728399e-02  -1.41487150e-02  -6.44477497e-02
-9.62241791e-04  -6.04519090e-02  -2.13903088e-01  -4.23907662e-04
 6.61842575e-02  -9.27999267e-02  -8.96045175e-02  -4.14201469e-02
 5.43293569e-02  -2.44044487e-02   7.74426959e-02   1.78279976e-01
-1.28434886e-02  -4.88005027e-02   2.80722202e-01  -8.38722651e-02
 4.52502022e-02  -3.53279475e-02   2.26220198e-02  -1.06247817e-01
-6.71949302e-02   7.33608894e-02  -5.56646882e-02  -2.75243021e-01
 9.42312124e-02   7.61649568e-02   5.23221056e-03   4.68681043e-02
-5.25306857e-02   5.69087867e-02   8.30759016e-02   2.35230057e-01
 1.32276695e-01  -3.29676945e-02   7.56616959e-03   5.52284895e-02
-3.27899178e-02  -1.65884257e-01  -4.83711896e-02   1.17233391e-01
 5.64983749e-02   1.85395090e-01   7.23360956e-02  -8.05226260e-02
 6.36369193e-02   3.62183185e-02  -1.87578951e-01   6.42483152e-02
```

```
 2.10763652e-01   1.44042220e-03   2.77749094e-02   3.45294387e-02
 9.55857249e-03   7.19821614e-03   3.75374322e-02  -2.65161263e-01
 7.76656261e-02   9.51133419e-02   6.92486049e-02  -5.96208099e-03
 2.03850041e-02   2.10154563e-01   6.32229951e-02  -1.01937132e-01
 1.44852737e-02   1.18637332e-01  -1.00933504e-03  -3.24656281e-02
-1.89712252e-02   2.35502025e-02   2.05906021e-02  -4.17720832e-02
 3.83917943e-02   3.05460351e-03   1.15549496e-02  -6.60407494e-02
-3.18256206e-02   2.71953405e-02  -1.41925174e-01  -1.20810227e-02
-3.86817356e-02   3.30279899e-03  -2.80285605e-02  -4.67068042e-02
-7.91255724e-02  -6.02026255e-03   3.38146050e-02   2.03458852e-02
 2.14964864e-02  -4.68431297e-03   4.38149505e-02   1.07008099e-01
-2.03066188e-01  -1.44966692e-01   1.32355842e-02   6.66614281e-03
 2.05912615e-02  -4.03710449e-02  -8.45746609e-02  -1.29474782e-03
-7.41308374e-02  -2.13159042e-02  -7.25053970e-03   1.61066217e-01
-1.27561226e-02   1.21893074e-01  -2.30583867e-02  -2.19971900e-02
-4.98237581e-02   2.63080852e-02  -6.18535672e-02   3.87696685e-02
-2.72733548e-01  -1.56310602e-02   2.09966980e-02  -2.16693653e-01
-8.02038834e-02   1.01795479e-01  -1.00259268e-01   1.07403085e-01
-1.11368472e-01  -1.55326828e-02   2.85036877e-01  -6.16632213e-02
 2.30366319e-02   5.62998070e-03   4.51701687e-02  -1.71996562e-01
-6.43556993e-03  -5.36092574e-02   1.78125217e-01   2.68392708e-02
-1.39581498e-02  -6.46418189e-02   2.33672303e-02  -2.50062038e-01
-1.01619743e-02  -9.13277215e-02  -1.98494402e-02  -6.23487408e-04
-1.78071450e-02   6.79290976e-02  -2.00519838e-01   1.83903292e-01
-3.30469481e-02   4.83270452e-02  -2.13862303e-01  -8.54764843e-02
 8.17735885e-02  -6.38663007e-02   1.05880928e-02  -1.27762093e-01
-1.10627868e-01  -1.32656899e-01  -1.24115963e-02  -3.18400340e-02
-5.02028273e-03   1.37823625e-01  -7.06180113e-02  -1.78076866e-02
-1.13155194e-03   6.14934101e-02  -5.44525044e-03   1.00491179e-01
 2.64591838e-01   1.18498713e-01  -3.50204724e-03  -8.93261649e-02
 4.27599707e-02   6.44563862e-02  -1.68713189e-02   1.39954012e-02
 8.92520013e-02  -1.99857954e-02   1.79582018e-02   1.03379333e-02
-5.42680139e-02  -3.39467252e-02   1.39173033e-01  -1.82993380e-02
-9.82854914e-02  -9.52523801e-03   9.31787408e-03  -8.50249973e-02
-3.55627199e-02  -6.34093458e-02   9.10443810e-03   2.29730843e-02
-7.19861240e-02  -4.35843745e-02   2.35518767e-02   1.23495123e-01
-6.26176048e-02   7.60280670e-02  -7.30598410e-03   2.42757627e-02
 2.46077587e-02  -5.59602070e-02   9.21090493e-02   9.58411534e-03
-1.69025182e-02   1.15292952e-01  -8.48786095e-02  -5.08668811e-02]]
```

In [102]:
```
#http://scikit-learn.org/stable/auto_examples/feature_selection/plot_rfe_with_cross_
from sklearn.model_selection import StratifiedKFold
from sklearn.feature_selection import RFECV
from sklearn.datasets import make_classification

# Create the RFE object and compute a cross-validated score.
```

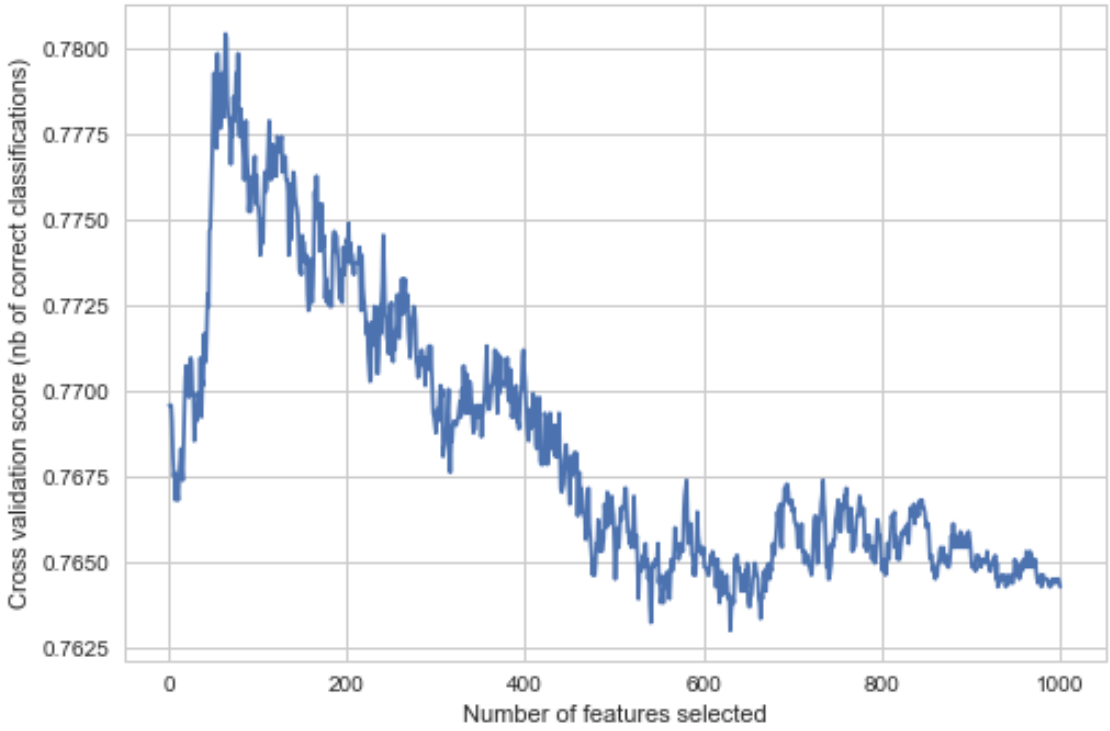
```
        # The "accuracy" scoring is proportional to the number of correct
        # classifications
        rfecv = RFECV(estimator=LogReg, step=1, cv=StratifiedKFold(10),
                      scoring='accuracy')
        rfecv.fit(X, y)

        print("Optimal number of features : %d" % rfecv.n_features_)
```

```
Optimal number of features : 64
```

In [ ]:

In [103]:
```
        # Plot number of features VS. cross-validation scores
        plt.figure()
        plt.xlabel("Number of features selected")
        plt.ylabel("Cross validation score (nb of correct classifications)")
        plt.plot(range(1, len(rfecv.grid_scores_) + 1), rfecv.grid_scores_)
        plt.show()
```



In [104]:
```
        logit_model=sm.Logit(y,rfecv.transform(X))
        result=logit_model.fit()
        print(result.summary())
```

```
Optimization terminated successfully.
         Current function value: 0.488907
         Iterations 6
                           Logit Regression Results
==============================================================================
Dep. Variable:            extroverted   No. Observations:                 8675
Model:                          Logit   Df Residuals:                     8611
Method:                           MLE   Df Model:                           63
Date:                Mon, 30 Apr 2018   Pseudo R-squ.:                0.09428
Time:                        01:25:08   Log-Likelihood:                -4241.3
converged:                       True   LL-Null:                       -4682.7
                                        LLR p-value:                6.310e-145
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
x1            -0.2796      0.082     -3.409      0.001      -0.440      -0.119
x2            -0.4072      0.087     -4.685      0.000      -0.578      -0.237
x3            -0.3751      0.080     -4.679      0.000      -0.532      -0.218
x4             0.0964      0.034      2.821      0.005       0.029       0.163
x5             0.2167      0.053      4.112      0.000       0.113       0.320
x6            -0.3678      0.086     -4.258      0.000      -0.537      -0.199
x7             0.2101      0.061      3.450      0.001       0.091       0.329
x8            -0.3686      0.085     -4.331      0.000      -0.535      -0.202
x9            -0.3287      0.067     -4.915      0.000      -0.460      -0.198
x10           -0.3497      0.066     -5.303      0.000      -0.479      -0.220
x11            0.1480      0.065      2.280      0.023       0.021       0.275
x12           -0.3967      0.094     -4.223      0.000      -0.581      -0.213
x13           -0.3960      0.063     -6.256      0.000      -0.520      -0.272
x14            0.2166      0.063      3.436      0.001       0.093       0.340
x15           -0.3510      0.085     -4.143      0.000      -0.517      -0.185
x16            0.1404      0.066      2.116      0.034       0.010       0.270
x17           -0.3618      0.073     -4.946      0.000      -0.505      -0.218
x18           -0.3571      0.089     -4.005      0.000      -0.532      -0.182
x19           -0.3590      0.093     -3.879      0.000      -0.540      -0.178
x20           -0.3880      0.057     -6.767      0.000      -0.500      -0.276
x21           -0.4044      0.083     -4.875      0.000      -0.567      -0.242
x22           -0.4773      0.089     -5.351      0.000      -0.652      -0.303
x23            0.1700      0.023      7.545      0.000       0.126       0.214
x24            0.1203      0.072      1.681      0.093      -0.020       0.261
x25            0.1447      0.077      1.882      0.060      -0.006       0.295
x26           -0.3226      0.039     -8.346      0.000      -0.398      -0.247
x27            0.0697      0.067      1.037      0.300      -0.062       0.202
x28            0.1176      0.079      1.494      0.135      -0.037       0.272
x29            0.1237      0.028      4.380      0.000       0.068       0.179
x30           -0.3085      0.053     -5.781      0.000      -0.413      -0.204
x31           -0.3753      0.078     -4.827      0.000      -0.528      -0.223
x32            0.0636      0.077      0.824      0.410      -0.088       0.215
x33           -0.3482      0.086     -4.065      0.000      -0.516      -0.180
```

```
x34            0.1757      0.051      3.441      0.001      0.076      0.276
x35            0.1278      0.061      2.092      0.036      0.008      0.247
x36           -0.3251      0.060     -5.430      0.000     -0.442     -0.208
x37            0.0249      0.064      0.387      0.699     -0.101      0.151
x38           -0.4431      0.064     -6.899      0.000     -0.569     -0.317
x39            0.0856      0.079      1.077      0.281     -0.070      0.241
x40            0.0914      0.070      1.311      0.190     -0.045      0.228
x41            0.0864      0.073      1.182      0.237     -0.057      0.230
x42           -0.4092      0.063     -6.537      0.000     -0.532     -0.287
x43            0.2524      0.025     10.295      0.000      0.204      0.300
x44           -0.2935      0.030     -9.751      0.000     -0.353     -0.235
x45            0.1028      0.058      1.772      0.076     -0.011      0.216
x46            0.0927      0.063      1.472      0.141     -0.031      0.216
x47           -0.3401      0.070     -4.877      0.000     -0.477     -0.203
x48           -0.3800      0.093     -4.104      0.000     -0.561     -0.199
x49           -0.3784      0.079     -4.818      0.000     -0.532     -0.224
x50           -0.3356      0.086     -3.892      0.000     -0.505     -0.167
x51           -0.4486      0.065     -6.866      0.000     -0.577     -0.321
x52           -0.3551      0.073     -4.851      0.000     -0.499     -0.212
x53           -0.3911      0.090     -4.350      0.000     -0.567     -0.215
x54           -0.3786      0.079     -4.768      0.000     -0.534     -0.223
x55            0.1229      0.079      1.561      0.119     -0.031      0.277
x56            0.1349      0.046      2.963      0.003      0.046      0.224
x57            0.0607      0.056      1.084      0.279     -0.049      0.170
x58           -0.4425      0.096     -4.610      0.000     -0.631     -0.254
x59           -0.3022      0.074     -4.059      0.000     -0.448     -0.156
x60           -0.3740      0.084     -4.446      0.000     -0.539     -0.209
x61            0.1402      0.058      2.406      0.016      0.026      0.254
x62           -0.3147      0.069     -4.567      0.000     -0.450     -0.180
x63            0.0515      0.076      0.673      0.501     -0.098      0.201
x64            0.0571      0.068      0.841      0.400     -0.076      0.190
==============================================================================
```

```
In [106]: rfecv.get_support()
```

```
Out[106]: array([False, False, False, False, False, False, False, False, False,
                 False, False, False, False, False, False, False, False, False,
                 False, False, False, False, False, False, False,  True,  True,
                 False, False, False, False, False, False, False, False, False,
                 False, False, False, False, False, False, False, False, False,
                 False, False, False, False, False,  True, False, False, False,
                  True, False, False, False, False, False, False, False, False,
                 False, False, False, False, False, False, False, False, False,
                 False, False, False, False,  True, False, False, False,  True,
                 False, False, False, False, False, False, False,  True, False,
                 False, False, False, False, False, False, False, False, False,
                 False,  True, False, False, False, False, False, False, False,
```

```
False, False, False, False, False, False, False, False, False,
False, False, False, False, False, False, False, False, False,
False, False, False,  True, False, False, False, False, False,
False, False, False,  True, False, False, False, False, False,
False, False, False, False, False, False, False, False, False,
False, False, False, False, False, False, False, False, False,
False, False, False, False, False, False, False, False, False,
False, False, False, False,  True, False, False, False, False,
False,  True, False,  True,  True, False, False, False, False,
False, False, False, False, False, False, False, False, False,
False, False, False, False,  True,  True, False, False, False,
False, False, False, False, False,  True, False, False, False,
False, False, False, False, False,  True, False, False,  True,
 True, False, False, False, False,  True, False, False, False,
 True, False, False, False, False, False, False, False, False,
False, False, False, False, False, False, False, False, False,
False, False, False, False, False, False, False,  True, False,
False, False, False,  True, False, False,  True, False, False,
False, False, False, False, False, False, False, False, False,
False, False, False, False, False, False, False,  True, False,
False, False, False, False, False, False, False, False, False,
False, False, False, False, False, False, False, False, False,
False, False,  True, False, False, False, False, False, False,
 True, False, False, False, False, False,  True, False,  True,
False, False, False, False, False, False, False, False, False,
False, False, False,  True, False, False, False, False, False,
False, False, False, False,  True, False, False, False, False,
False, False, False, False,  True, False, False, False, False,
False, False, False, False, False, False, False, False, False,
False, False, False, False, False, False, False, False, False,
False, False, False, False, False, False, False, False, False,
False, False, False, False, False, False, False, False,  True,
False, False, False, False, False, False, False, False, False,
False, False, False, False, False, False, False, False, False,
False, False, False, False, False, False, False,  True, False,
False, False, False, False, False, False, False, False, False,
False, False, False, False, False, False, False, False, False,
False, False, False, False, False, False, False, False, False,
False, False, False, False, False, False, False, False, False,
False, False, False, False, False, False, False, False, False,
False, False, False, False, False, False, False, False, False,
False, False,  True, False, False, False,  True, False, False,
False, False, False, False, False, False, False, False, False,
False, False, False, False, False, False, False, False, False,
False, False, False, False,  True, False,  True, False, False,
```

```
False, False, False, False, False, False, False, False,  True,
False, False, False, False, False, False, False, False, False,
False, False, False, False, False, False, False, False, False,
False, False, False, False, False, False, False, False, False,
False, False, False, False, False, False, False, False, False,
 True,  True, False, False, False, False, False, False, False,
False, False, False,  True, False, False, False, False, False,
False, False, False, False,  True, False, False, False, False,
False, False, False, False, False,  True, False, False, False,
False, False, False, False, False, False, False, False, False,
False, False,  True, False,  True, False, False, False, False,
False, False, False, False, False, False, False, False, False,
 True, False, False, False, False, False, False, False, False,
False,  True, False, False, False, False, False, False, False,
False, False, False, False, False, False, False, False, False,
False, False, False, False, False, False, False, False, False,
False, False, False, False, False, False, False, False, False,
False,  True, False, False, False, False,  True, False, False,
False, False, False, False, False, False, False, False, False,
False, False, False, False, False, False, False, False, False,
False, False, False, False, False, False, False, False, False,
False, False, False, False, False, False, False, False, False,
False, False, False, False, False, False, False, False, False,
False, False, False, False, False, False, False, False, False,
False, False, False, False,  True, False, False, False, False,
False, False, False, False, False, False, False, False, False,
 True, False, False, False, False, False, False, False, False,
False, False, False, False, False, False, False, False, False,
False, False, False, False, False, False, False, False, False,
 True, False, False, False, False, False, False, False, False,
False, False, False, False, False, False, False, False, False,
False, False,  True, False, False, False, False, False, False,
False, False, False, False, False, False, False, False, False,
False, False, False, False, False, False, False, False, False,
False, False, False, False, False, False, False, False, False,
False, False, False, False, False, False, False, False,  True,
False, False, False, False, False, False, False, False, False,
False, False, False, False, False, False, False, False, False,
False, False, False, False,  True, False, False, False, False,
False, False,  True,  True, False, False, False, False, False,
False, False,  True, False, False, False, False, False, False,
False, False, False, False, False, False,  True,  True, False,
False, False, False, False, False, False, False, False, False,
```

```
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False,
       False, False, False, False, False, False, False, False, False, False], dtype=b
```

In [120]:
```
support = rfecv.get_support()
#Now support is an array, you can use that to efficiently extract the name of your s

feature_names = np.array(count_vectorizer.get_feature_names()) # transformed list to

feature_names_support = feature_names[support]
```

In [108]:
```
rfecv.score(X, y)
```

Out[108]:
```
0.7887031700288184
```

In [146]:
```
#change max rows to show
pd.options.display.max_rows =100
pd.options.display.max_rows
```

Out[146]:
```
100
```

In [148]:
```
final = pd.DataFrame({'feature name':feature_names_support, 'coefficient': result.par
#https://stackoverflow.com/questions/13148429/how-to-change-the-order-of-dataframe-c
def order(frame,var):
    if type(var) is str:
        var = [var] #let the command take a string or list
    varlist =[w for w in frame.columns if w not in var]
    frame = frame[var+varlist]
    return frame
final = order(final, ['feature name'])
final = final.sort_values('coefficient', ascending = False)
final
```

Out[148]:

	feature name	coefficient
x43	ne	0.252407
x5	bored	0.216664
x14	debate	0.216558
x7	business	0.210116
x34	hahaha	0.175742
x23	estp	0.170026
x11	dated	0.148015
x25	excited	0.144750
x16	developed	0.140425
x61	wanna	0.140193
x56	super	0.134935
x35	hot	0.127757
x29	fun	0.123675
x55	spot	0.122939
x24	exact	0.120324

x28	friendly	0.117644
x45	nt	0.102775
x4	awesome	0.096438
x46	omg	0.092722
x40	mad	0.091433
x41	match	0.086433
x39	loves	0.085567
x27	fight	0.069748
x32	goal	0.063616
x57	term	0.060655
x64	win	0.057083
x63	willing	0.051451
x37	line	0.024871
x1	animals	-0.279573
x44	ni	-0.293545
x59	videos	-0.302228
x30	games	-0.308529
x62	watched	-0.314680
x26	family	-0.322591
x36	inferior	-0.325118
x9	comfortable	-0.328738
x50	played	-0.335559
x47	particularly	-0.340056
x33	green	-0.348198
x10	computer	-0.349727
x15	details	-0.351033
x52	rarely	-0.355133
x18	door	-0.357116
x19	draw	-0.358962
x17	dislike	-0.361827
x6	boy	-0.367780
x8	cats	-0.368608
x60	walking	-0.374019
x3	average	-0.375111
x31	gender	-0.375306
x49	philosophy	-0.378426
x54	soul	-0.378604
x48	parts	-0.379960
x20	dream	-0.387978
x53	sign	-0.391099
x13	death	-0.396006
x12	dealing	-0.396716
x21	dry	-0.404414
x2	anime	-0.407223
x42	nature	-0.409219
x58	uncomfortable	-0.442473
x38	listening	-0.443149
x51	quiet	-0.448619

```
x22           earth     -0.477344
```