```
In [31]:   import pandas as pd
           import numpy as np
```

```
In [32]:   df1=pd.read_csv("C:/Users/SANDEEP/Desktop/skype photos&videos/spam1.csv")
```

```
In [33]:   df1
```

Out[33]:

	Category	Message
0	ham	Go until jurong point, crazy.. Available only ...
1	ham	Ok lar... Joking wif u oni...
2	spam	Free entry in 2 a wkly comp to win FA Cup fina...
3	ham	U dun say so early hor... U c already then say...
4	ham	Nah I don't think he goes to usf, he lives aro...
...	...	...
5567	spam	This is the 2nd time we have tried 2 contact u...
5568	ham	Will ü b going to esplanade fr home?
5569	ham	Pity, * was in mood for that. So...any other s...
5570	ham	The guy did some bitching but I acted like i'd...
5571	ham	Rofl. Its true to its name

5572 rows × 2 columns

```
In [34]:   df1["Mail"]=df1["Category"].apply(lambda x:1 if x=="spam" else 0)
```

```
In [35]:   df1
```

	Category	Message	Mail
0	ham	Go until jurong point, crazy.. Available only ...	0
1	ham	Ok lar... Joking wif u oni...	0
2	spam	Free entry in 2 a wkly comp to win FA Cup fina...	1
3	ham	U dun say so early hor... U c already then say...	0
4	ham	Nah I don't think he goes to usf, he lives aro...	0
...	...	...	...
5567	spam	This is the 2nd time we have tried 2 contact u...	1
5568	ham	Will ü b going to esplanade fr home?	0
5569	ham	Pity, * was in mood for that. So...any other s...	0
5570	ham	The guy did some bitching but I acted like i'd...	0
5571	ham	Rofl. Its true to its name	0

5572 rows × 3 columns

In [36]:
```python
df1.drop("Category",axis=1,inplace=True)
```

In [37]:
```python
df1
```

Out[37]:

	Message	Mail
0	Go until jurong point, crazy.. Available only ...	0
1	Ok lar... Joking wif u oni...	0
2	Free entry in 2 a wkly comp to win FA Cup fina...	1
3	U dun say so early hor... U c already then say...	0
4	Nah I don't think he goes to usf, he lives aro...	0
...	...	...

	Message	Mail
5567	This is the 2nd time we have tried 2 contact u...	1
5568	Will ü b going to esplanade fr home?	0
5569	Pity, * was in mood for that. So...any other s...	0
5570	The guy did some bitching but I acted like i'd...	0
5571	Rofl. Its true to its name	0

5572 rows × 2 columns

for converting msg to machine learning i.e numaric we have to follow some steps

1)tokenize:it will splitting the data into words

2)identify the unique words then after

3)give the indentification number each and every unique words

tha above three steps done by only one single functions that is vectorizer vectorizer has classified two types

1)count vectorizer

2)tfidf vectorizer

Mostly we can use count verctorizer in this machine learning and tfidf we can use in deep learning the count vectorizer do the above all three steps at a time just we have to import that is and fit it.It will take care of every thing

```
In [38]: from sklearn.model_selection import train_test_split
```

```
In [40]: Xtrain,Xtest,ytrain,ytest=train_test_split(df1.Message,df1.Mail,test_size=0.20)
```

```
In [41]: Xtrain.shape
```

```
Out[41]: (4457,)
```

```
In [42]: ytrain.shape
```

```
Out[42]: (4457,)
```

```
In [43]: Xtest.shape
```

```
Out[43]: (1115,)
```

```
In [44]: ytest.shape
```

```
Out[44]: (1115,)
```

```
In [45]: from sklearn.feature_extraction.text import CountVectorizer
```

```
In [46]: cv=CountVectorizer()
```

```
In [55]: cv.fit(Xtrain.values)
```

```
Out[55]: CountVectorizer()
```

```
In [56]: cv.get_feature_names()
```

```
Out[56]: ['00',
         '000',
         '000pes',
         '008704050406',
         '0089',
         '0121',
         '01223585334',
         '0125698789',
```

```
'02',
'0207',
'02073162414',
'02085076972',
'021',
'03',
'04',
'0430',
'05',
'050703',
'0578',
'06',
'07',
'07008009200',
'07046744435',
'07090201529',
'07090298926',
'07099833605',
'07123456789',
'0721072',
'07734396839',
'07742676969',
'07753741225',
'0776xxxxxxx',
'07781482378',
'07786200117',
'077xxx',
'078',
'07801543489',
'07808',
'07808726822',
'07815296484',
'07821230901',
'078498',
'07880867867',
'0789xxxxxxx',
'07946746291',
'0796xxxxxx',
'07973788240',
'07xxxxxxxxx',
'08',
'0800',
'08000407165',
'08000776320',
'08000839402',
```

```
'08000930705',
'08001950382',
'08002888812',
'08002986030',
'08002986906',
'08006344447',
'0808',
'08081263000',
'08081560665',
'0825',
'083',
'0844',
'08448350055',
'08448714184',
'0845',
'08450542832',
'08452810073',
'08452810075over18',
'0870',
'08700435505150p',
'08700469649',
'08700621170150p',
'08701213186',
'08701237397',
'08701417012',
'08701417012150p',
'0870141701216',
'087016248',
'08701752560',
'087018728737',
'0870241182716',
'08702490080',
'08702840625',
'08704050406',
'08704439680',
'08706091795',
'0870737910216yrs',
'08707500020',
'08707509020',
'0870753331018',
'08708034412',
'08708800282',
'08709222922',
'0871',
'087104711148',
```

```
'08712101358',
'08712103738',
'0871212025016',
'08712300220',
'087123002209am',
'08712317606',
'08712400200',
'08712400602450p',
'08712400603',
'08712402050',
'08712402578',
'08712402902',
'08712402972',
'08712404000',
'08712405020',
'08712405022',
'08712460324',
'08712466669',
'0871277810810',
'0871277810910p',
'08714342399',
'087147123779am',
'08714712379',
'08714712388',
'08714712394',
'08714712412',
'08714714011',
'08715203028',
'08715203649',
'08715203652',
'08715203656',
'08715203677',
'08715203685',
'08715203694',
'08715205273',
'08715500022',
'08715705022',
'08717111821',
'0871750',
'08717507382',
'08717509990',
'08717890890',
'08717898035',
'08718711108',
'08718720201',
```

```
'08718725756',
'08718726270',
'087187262701',
'08718726971',
'08718726978',
'08718727870',
'08718727870150ppm',
'08718730555',
'08718730666',
'08718738001',
'08718738002',
'08719180219',
'08719180248',
'08719181259',
'08719181503',
'08719181513',
'08719839835',
'08719899217',
'09',
'09041940223',
'09050000301',
'09050000332',
'09050000460',
'09050000555',
'09050000878',
'09050001295',
'09050001808',
'09050003091',
'09050090044',
'09050280520',
'09053750005',
'09056242159',
'09058091854',
'09058091870',
'09058094454',
'09058094455',
'09058094507',
'09058094565',
'09058094594',
'09058094597',
'09058094599',
'09058095107',
'09058095201',
'09058097189',
'09058099801',
```

```
'09061104276',
'09061104283',
'09061209465',
'09061213237',
'09061221061',
'09061221066',
'09061701444',
'09061701461',
'09061701851',
'09061701939',
'09061702893',
'09061743386',
'09061743806',
'09061743811',
'09061744553',
'09061749602',
'09061790121',
'09061790125',
'09063440451',
'09063442151',
'09063458130',
'0906346330',
'09064011000',
'09064012160',
'09064015307',
'09064017295',
'09064017305',
'09064018838',
'09064019014',
'09064019788',
'09065069120',
'09065069154',
'09065171142',
'09065174042',
'09065394514',
'09065989180',
'09065989182',
'09066350750',
'09066358152',
'09066358361',
'09066361921',
'09066362206',
'09066362220',
'09066362231',
'09066364311',
```

```
'09066364349',
'09066364589',
'09066368327',
'09066368753',
'09066380611',
'09066382422',
'09066612661',
'09066660100',
'09071512432',
'09071512433',
'09071517866',
'09077818151',
'09090204448',
'09090900040',
'09094100151',
'09094646631',
'09094646899',
'09095350301',
'09096102316',
'09099725823',
'09099726395',
'09099726481',
'09099726553',
'09111030116',
'09111032124',
'0a',
'0quit',
'10',
'100',
'1000',
'1000call',
'1000s',
'100p',
'100percent',
'100txt',
'1013',
'1030',
'10am',
'10k',
'10p',
'10ppm',
'10th',
'11',
'1120',
'113',
```

```
'1131',
'114',
'1146',
'116',
'1172',
'11mths',
'11pm',
'12',
'1205',
'120p',
'121',
'1225',
'123',
'1250',
'125gift',
'12hours',
'12hrs',
'12mths',
'13',
'1327',
'14',
'140',
'1405',
'140ppm',
'145',
'1450',
'146tf150p',
'14tcr',
'14thmarch',
'15',
'150',
'1500',
'150p',
'150p16',
'150pm',
'150ppermesssubscription',
'150ppm',
'150ppmpobox10183bhamb64xe',
'150ppmsg',
'150pw',
'151',
'153',
'15541',
'15pm',
'16',
```

```
'165',
'1680',
'169',
'177',
'18',
'180',
'1843',
'18p',
'18yrs',
'1956669',
'1apple',
'1b6a5ecef91ff9',
'1cup',
'1da',
'1er',
'1hr',
'1lemon',
'1million',
'1st',
'1st4terms',
'1stone',
'1thing',
'1tulsi',
'1win150ppmx3',
'1winaweek',
'1winawk',
'1x150p',
'1yf',
'20',
'200',
'2000',
'2003',
'2004',
'2005',
'2006',
'2007',
'200p',
'2025050',
'20m12aq',
'20p',
'21',
'21870000',
'21st',
'22',
'220',
```

```
'220cm2',
'2309',
'23f',
'23g',
'24',
'24m',
'24th',
'25',
'250',
'250k',
'25p',
'26',
'2667',
'26th',
'27',
'28',
'2814032',
'28days',
'28th',
'28thfeb',
'29',
'2bold',
'2c',
'2day',
'2docd',
'2end',
'2ez',
'2find',
'2getha',
'2geva',
'2go',
'2gthr',
'2hrs',
'2i',
'2kbsubject',
'2lands',
'2marrow',
'2moro',
'2morow',
'2morro',
'2morrow',
'2morrowxxxx',
'2mro',
'2mrw',
'2mwen',
```

```
'2nd',
'2nhite',
'2nite',
'2optout',
'2p',
'2price',
'2px',
'2rcv',
'2stop',
'2stoptx',
'2stoptxt',
'2u',
'2waxsto',
'2wks',
'2wt',
'2years',
'2yr',
'2yrs',
'30',
'300',
'3000',
'300603',
'300603t',
'300p',
'3030',
'30apr',
'30ish',
'30pm',
'30pp',
'30s',
'31',
'3100',
'310303',
'31p',
'32',
'32000',
'3230',
'32323',
'326',
'33',
'330',
'350',
'3510i',
'35p',
'3650',
```

```
'36504',
'3680',
'373',
'3750',
'37819',
'38',
'382',
'3aj',
'3d',
'3days',
'3g',
'3gbp',
'3hrs',
'3lions',
'3lp',
'3miles',
'3mins',
'3mobile',
'3pound',
'3qxj9',
'3rd',
'3ss',
'3uz',
'3wks',
'3x',
'3xx',
'40',
'400',
'400mins',
'400thousad',
'402',
'4041',
'40533',
'40gb',
'41685',
'41782',
'420',
'42049',
'42478',
'42810',
'430',
'434',
'440',
'4403ldnw1a7rw18',
'44345',
```

```
'447797706009',
'447801259231',
'448712404000',
'449050000301',
'45',
'450',
'450p',
'450ppw',
'450pw',
'45239',
'45pm',
'47',
'4742',
'47per',
'48',
'4882',
'48922',
'49',
'49557',
'4a',
'4brekkie',
'4d',
'4eva',
'4few',
'4fil',
'4get',
'4give',
'4got',
'4goten',
'4info',
'4jx',
'4msgs',
'4my',
'4qf2',
'4t',
'4th',
'4thnov',
'4txt',
'4u',
'4utxt',
'4w',
'4wrd',
'4xx26',
'4years',
'50',
```

```
'500',
'5000',
'505060',
'50award',
'50ea',
'50gbp',
'50p',
'50perweeksub',
'50perwksub',
'50pm',
'50pmmorefrommobile2bremoved',
'50ppm',
'50rcvd',
'50s',
'515',
'5226',
'5249',
'528',
'530',
'542',
'545',
'5free',
'5ish',
'5k',
'5min',
'5mls',
'5p',
'5pm',
'5th',
'5wb',
'5we',
'5wkg',
'5wq',
'5years',
'60',
'600',
'6031',
'6089',
'60p',
'61',
'61200',
'61610',
'62220cncl',
'6230',
'62468',
```

```
'62735',
'630',
'645',
'65',
'650',
'66',
'674',
'67441233',
'68866',
'69101',
'69669',
'69696',
'69698',
'69866',
'69876',
'69888',
'69888nyt',
'69911',
'69969',
'69988',
'6days',
'6hl',
'6hrs',
'6ish',
'6months',
'6ph',
'6pm',
'6th',
'6times',
'6wu',
'700',
'7250',
'7250i',
'730',
'731',
'74355',
'75',
'750',
'7548',
'75max',
'762',
'77',
'78',
'786',
'7876150ppm',
```

'7am',
'7cfca1a',
'7ish',
'7mp',
'7pm',
'7th',
'7ws',
'7zs',
'800',
'8000930705',
'80062',
'8007',
'80082',
'80086',
'80122300p',
'80155',
'80160',
'80182',
'8027',
'80488',
'80608',
'8077',
'80878',
'81010',
'81151',
'81303',
'81618',
'82050',
'820554ad0a1705572711',
'82242',
'82277',
'82324',
'82468',
'83021',
'83039',
'83049',
'83110',
'83118',
'83222',
'83332',
'83338',
'83355',
'83370',
'83383',
'83600',

```
'83738',
'84025',
'84122',
'84128',
'84199',
'85',
'850',
'85023',
'85069',
'85233',
'8552',
'85555',
'86021',
'861',
'864233',
'86688',
'86888',
'87021',
'87066',
'87070',
'87077',
'87121',
'87131',
'8714714',
'872',
'87239',
'87575',
'8800',
'88039',
'88066',
'88088',
'88222',
'88600',
'88800',
'88888',
'89034',
'89070',
'89080',
'89105',
'89123',
'89545',
'89555',
'89693',
'89938',
'8am',
```

```
'8ball',
'8p',
'8pm',
'8th',
'8wp',
'900',
'9061100010',
'910',
'9280114',
'92h',
'930',
'9307622',
'945',
'946',
'95',
'9755',
'9758',
'97n7qp',
'98321561',
'99',
'9996',
'9ae',
'9am',
'9ja',
'9pm',
'9t',
'9th',
'9yt',
'‾‾‾',
'a21',
'aa',
'aah',
'aaniye',
'aaooooright',
'aathi',
'ab',
'abbey',
'abel',
'abi',
'ability',
'abiola',
'abj',
'able',
'abnormally',
'about',
```

```
'aboutas',
'above',
'abroad',
'absolutely',
'absolutly',
'abstract',
'abt',
'abta',
'aburo',
'abuse',
'abusers',
'ac',
'academic',
'acc',
'accent',
'accenture',
'accept',
'access',
'accident',
'accidentally',
'accommodation',
'accommodationvouchers',
'accomodations',
'accordingly',
'account',
'accounting',
'achan',
'ache',
'achieve',
'acknowledgement',
'acl03530150pm',
'acnt',
'aco',
'across',
'act',
'acted',
'actin',
'acting',
'action',
'activ8',
'activate',
'active',
'activities',
'actor',
'actual',
```

```
'actually',
'ad',
'adam',
'add',
'addamsfa',
'added',
'addicted',
'addie',
'adding',
'address',
'adds',
'adewale',
'adjustable',
'admin',
'administrator',
'admirer',
'admission',
'admit',
'adore',
'adoring',
'adress',
'adrian',
'ads',
'adsense',
'adult',
'adults',
'advance',
'adventure',
'adventuring',
'advice',
'advise',
'advising',
'advisors',
'aeronautics',
'aeroplane',
'afew',
'affair',
'affection',
'affectionate',
'affections',
'affidavit',
'afford',
'afghanistan',
'afraid',
'africa',
```

```
'african',
'aft',
'after',
'afternon',
'afternoon',
'afternoons',
'aftr',
'ag',
'again',
'against',
'agalla',
'age',
'age16',
'age23',
'agency',
'agent',
'agents',
'ages',
'agidhane',
'aging',
'ago',
'agree',
'ah',
'aha',
'ahead',
'ahhh',
'ahhhh',
'ahmad',
'aid',
'aids',
'aig',
'aight',
'ain',
'aint',
'air',
'air1',
'airport',
'airtel',
'aiya',
'aiyah',
'aiyar',
'aiyo',
'ak',
'aka',
'akon',
```

```
'al',
'alaikkum',
'alaipayuthe',
'albi',
'album',
'alcohol',
'aldrine',
'alert',
'alertfrom',
'alerts',
'aletter',
'alex',
'alfie',
'algarve',
'algorithms',
'ali',
'alian',
'alive',
'all',
'allah',
'allalo',
'allday',
'alle',
'allo',
'allow',
'allowed',
'allows',
'almost',
'alone',
'along',
'alot',
'already',
'alright',
'alrite',
'also',
'alternative',
'although',
'alto18',
'aluable',
'always',
'alwys',
'am',
'amanda',
'amazing',
'ambitious',
```

```
'ambrith',
'american',
'ami',
'amigos',
'amla',
'amma',
'ammae',
'ammo',
'amnow',
'among',
'amongst',
'amore',
'amount',
'amp',
'amrca',
'amrita',
'ams',
'amt',
'amused',
'amy',
'an',
'ana',
'anal',
'and',
'anderson',
'andre',
'andres',
'andros',
'angels',
'angry',
'animal',
'animation',
'anjie',
'anjola',
'anna',
'annie',
'anniversary',
'annoncement',
'announced',
'announcement',
'annoyin',
'annoying',
'anonymous',
'anot',
'another',
```

```
        'ans',
        'ansr',
        ...]
```

In [57]: `cv.vocabulary_`

Out[57]:
```
{'sure': 6650,
 'but': 1590,
 'make': 4356,
 'he': 3395,
 'knows': 3994,
 'we': 7443,
 'ain': 895,
 'smokin': 6281,
 'yet': 7719,
 'cheers': 1776,
 'for': 2954,
 'callin': 1631,
 'babe': 1204,
 'sozi': 6369,
 'culdnt': 2124,
 'talkbut': 6723,
 'wannatell': 7402,
 'details': 2289,
 'later': 4061,
 'wenwecan': 7483,
 'chat': 1758,
 'properly': 5491,
 'watever': 7429,
 'relation': 5701,
 'built': 1570,
 'up': 7208,
 'in': 3678,
 'dis': 2361,
 'world': 7624,
 'only': 4973,
 'thing': 6866,
 'which': 7516,
 'remains': 5712,
 'atlast': 1138,
 'iz': 3800,
 'lonlines': 4223,
 'with': 7574,
 'lotz': 4250,
```

```
'lot': 4244,
'memories': 4469,
'feeling': 2823,
'stayin': 6482,
'here': 3442,
'an': 973,
'extra': 2750,
'week': 7462,
'back': 1211,
'next': 4797,
'wed': 7455,
'how': 3555,
'did': 2318,
'do': 2394,
'the': 6837,
'rugby': 5867,
'this': 6875,
'weekend': 7464,
'hi': 3455,
'to': 6957,
'and': 976,
'soon': 6342,
'email': 2588,
'alertfrom': 916,
'jeri': 3841,
'stewartsize': 6499,
'2kbsubject': 402,
'low': 4270,
'cost': 2032,
'prescripiton': 5415,
'drvgsto': 2488,
'listen': 4172,
'call': 1619,
'123': 290,
'good': 3203,
'evening': 2680,
'is': 3774,
'roger': 5834,
'are': 1067,
'you': 7732,
'job': 3859,
'profile': 5469,
'seems': 6014,
'like': 4141,
'bpo': 1489,
```

```
'uncle': 7164,
'abbey': 764,
'happy': 3360,
'new': 4789,
'year': 7707,
'abiola': 768,
'there': 6855,
'generally': 3121,
'isn': 3781,
'one': 4967,
'it': 3786,
'uncountable': 7170,
'noun': 4869,
'dictionary': 2317,
'pieces': 5238,
'of': 4921,
'research': 5752,
'want': 7403,
'lick': 4126,
'your': 7738,
'pussy': 5537,
'now': 4872,
'nvm': 4894,
'ok': 4944,
'called': 1626,
'said': 5901,
'all': 926,
'him': 3464,
'then': 6849,
'have': 3386,
'choose': 1815,
'future': 3067,
'short': 6140,
'cute': 2144,
'be': 1292,
'person': 5190,
'dont': 2427,
'try': 7083,
'prove': 5499,
'gud': 3287,
'noon': 4843,
'nice': 4800,
'day': 2194,
'today': 6963,
'love': 4257,
```

'so': 6304,
'dearly': 2208,
'please': 5279,
'our': 5035,
'customer': 2140,
'service': 6055,
'representative': 5745,
'on': 4961,
'0800': 49,
'169': 325,
'6031': 584,
'between': 1353,
'10am': 270,
'9pm': 752,
'as': 1101,
'won': 7603,
'guaranteed': 3285,
'1000': 262,
'cash': 1688,
'or': 5001,
'5000': 549,
'prize': 5449,
'lol': 4215,
'they': 6862,
'don': 2423,
'know': 3991,
'about': 772,
'my': 4693,
'awesome': 1188,
'phone': 5215,
'could': 2040,
'click': 1854,
'delete': 2246,
'right': 5806,
'if': 3634,
'sir': 6208,
'need': 4758,
'velusamy': 7292,
'date': 2186,
'birth': 1385,
'company': 1936,
'bank': 1238,
'facilities': 2762,
'me': 4434,
'da': 2155,

```
'am': 949,
'waiting': 7381,
'goodmorning': 3208,
'late': 4058,
'2hrs': 400,
'because': 1301,
'pain': 5081,
'hey': 3451,
'will': 7544,
'ah': 885,
'meet': 4451,
'at': 1133,
'945': 740,
'nope': 4846,
'watching': 7426,
'tv': 7108,
'home': 3507,
'not': 4860,
'going': 3191,
'out': 5038,
'bored': 1450,
'really': 5637,
'dhanush': 2303,
'rocks': 5832,
'once': 4965,
'again': 871,
'private': 5448,
'2003': 354,
'account': 797,
'statement': 6476,
'07815296484': 39,
'shows': 6163,
'800': 646,
'un': 7159,
'redeemed': 5676,
'points': 5316,
'08718738001': 152,
'identifier': 3628,
'code': 1890,
'41782': 493,
'expires': 2739,
'18': 327,
'11': 275,
'04': 14,
'very': 7299,
```

```
'environment': 2639,
'terrific': 6803,
'food': 2945,
'wen': 7481,
'miss': 4536,
'someone': 6322,
'definitely': 2236,
'special': 6383,
'why': 7529,
'them': 6846,
'just': 3904,
'keep': 3939,
'touch': 7023,
'gdeve': 3113,
'can': 1643,
'where': 7510,
'depends': 2268,
'wan': 7398,
'lor': 4236,
'busy': 1589,
'trying': 7084,
'finish': 2874,
'looking': 4229,
'forward': 2977,
'finally': 2866,
'meeting': 4453,
'squeezed': 6441,
'lt': 4279,
'gt': 3282,
'bucks': 1558,
'dad': 2158,
'dear': 2205,
'voucher': 7354,
'holder': 3499,
'claim': 1832,
'weeks': 7467,
'offer': 4926,
'pc': 5155,
'go': 3180,
'http': 3570,
'www': 7666,
'tlp': 6949,
'co': 1881,
'uk': 7148,
'reward': 5795,
```

'ts': 7085,
'cs': 2108,
'apply': 1041,
'find': 2868,
'what': 7499,
'time': 6925,
'bus': 1585,
'coz': 2059,
'sort': 6351,
'some': 6318,
'stuff': 6568,
'ambrith': 953,
'madurai': 4337,
'met': 4491,
'arun': 1100,
'dha': 2302,
'marrge': 4394,
'remembr': 5717,
'thats': 6836,
'cool': 2014,
'its': 3793,
'off': 4923,
'either': 2571,
'raglan': 5579,
'rd': 5619,
'edward': 2550,
'behind': 1324,
'cricket': 2090,
'ground': 3272,
'gimme': 3154,
'ring': 5812,
'when': 7506,
'ur': 7225,
'closeby': 1861,
'see': 6008,
'tuesday': 7095,
'december': 2212,
'had': 3315,
'mobile': 4569,
'11mths': 283,
'entitled': 2632,
'update': 7213,
'latest': 4062,
'colour': 1909,
'camera': 1639,

```
'free': 2996,
'vco': 7284,
'08002986906': 57,
'eatin': 2535,
'wif': 7535,
'frens': 3006,
'lei': 4103,
'first': 2886,
'whore': 7525,
'unbelievable': 7161,
'oh': 4939,
'after': 865,
'that': 6833,
'placement': 5258,
'jamster': 3814,
'get': 3135,
'wallpaper': 7393,
'text': 6812,
'heart': 3409,
'88888': 717,
'16': 322,
'help': 3428,
'08701213186': 75,
'checking': 1772,
'seeing': 6010,
'jeremiah': 3840,
'great': 3257,
'month': 4602,
'buy': 1594,
'space': 6371,
'invaders': 3749,
'chance': 1735,
'win': 7547,
'orig': 5017,
'arcade': 1064,
'game': 3084,
'console': 1988,
'press': 5423,
'games': 3085,
'std': 6485,
'wap': 7408,
'charge': 1746,
'o2': 4904,
'terms': 6799,
'settings': 6061,
```

'no': 4824,
'purchase': 5529,
'hero': 3443,
'leaving': 4093,
'qatar': 5545,
'tonite': 6996,
'apt': 1058,
'opportunity': 4992,
'pls': 5286,
'kerala': 3946,
'also': 942,
'thk': 6876,
'too': 6999,
'fast': 2799,
'xy': 7691,
'suggest': 6609,
'dun': 2505,
'rain': 5583,
'leh': 4102,
'got': 3221,
'gd': 3112,
'urgent': 7229,
'contact': 1991,
'todays': 6964,
'draw': 2458,
'2000': 353,
'09066358361': 227,
'from': 3032,
'land': 4035,
'line': 4155,
'y87': 7692,
'valid': 7269,
'12hrs': 294,
'willing': 7545,
'apps': 1053,
'class': 1840,
'guess': 3291,
'both': 1458,
'screwd': 5983,
'shuhui': 6171,
'say': 5948,
'change': 1737,
'suntec': 6632,
'steamboat': 6490,
'noe': 4828,

```
'eldest': 2578,
've': 7286,
'been': 1313,
'invited': 3753,
'xchat': 7674,
'final': 2864,
'attempt': 1143,
'txt': 7120,
'86688': 698,
'150p': 310,
'msgrcvdhg': 4647,
'suite342': 6613,
'2lands': 403,
'row': 5850,
'w1j6hl': 7363,
'ldn': 4080,
'yrs': 7750,
'cartoon': 1686,
'listening': 4175,
'music': 4683,
'amp': 966,
'eve': 2677,
'temple': 6790,
'church': 1825,
'days': 2195,
'billion': 1374,
'classes': 1841,
'continent': 1998,
'put': 5538,
'party': 5124,
'study': 6565,
'lightly': 4139,
'think': 6868,
'custom': 2139,
'checkboxes': 1769,
're': 5621,
'hardcore': 3362,
'girl': 3157,
'difficult': 2333,
'girls': 3160,
'friend': 3016,
'companion': 1935,
'lover': 4262,
'chef': 1781,
'listener': 4174,
```

```
'organizer': 5015,
'boyfriend': 1486,
'clean': 1846,
'sympathetic': 6692,
'athletic': 1135,
'warm': 7410,
'courageous': 2051,
'determined': 2291,
'true': 7075,
'dependable': 2267,
'intelligent': 3735,
'psychologist': 5511,
'pest': 5199,
'exterminator': 2749,
'psychiatrist': 5510,
'healer': 3403,
'stylist': 6575,
'driver': 2471,
'aaniye': 760,
'pudunga': 5517,
'venaam': 7293,
'yes': 7716,
'footie': 2952,
'worried': 7627,
'blow': 1418,
'phil': 5210,
'neville': 4787,
'mean': 4437,
'made': 4334,
'hold': 3498,
'weed': 7461,
'calls': 1633,
'afternoon': 867,
'onwords': 4979,
'mtnl': 4657,
'problem': 5456,
'south': 6366,
'mumbai': 4671,
'hear': 3405,
'cann': 1651,
'yalru': 7696,
'lyfu': 4303,
'astne': 1127,
'chikku': 1795,
'bt': 1555,
```

```
'innu': 3721,
'mundhe': 4675,
'lyf': 4302,
'ali': 923,
'halla': 3329,
'ke': 3937,
'bilo': 1377,
'marriage': 4395,
'program': 5472,
'edhae': 2544,
'nt': 4878,
'ovr': 5059,
'vargu': 7276,
'meow': 4478,
'wonders': 7609,
'7th': 643,
'6th': 620,
'style': 6572,
'5th': 576,
'smile': 6273,
'4th': 538,
'personality': 5193,
'3rd': 478,
'nature': 4738,
'2nd': 413,
'sms': 6284,
'1st': 341,
'lovely': 4260,
'friendship': 3019,
'morning': 4612,
'hiya': 3478,
'spoken': 6418,
'since': 6199,
'200': 352,
'shopping': 6137,
'spree': 6433,
'every': 2686,
'starting': 6470,
'play': 5270,
'store': 6525,
'88039': 711,
'skilgme': 6227,
'tscs08714740323': 7088,
'1winawk': 348,
'age16': 875,
```

'50perweeksub': 555,
'ya': 7693,
'thinking': 6871,
'down': 2445,
'reg': 5689,
'pract': 5382,
'lessons': 4115,
'flung': 2924,
'advance': 844,
'haha': 3317,
'wat': 7422,
'takes': 6717,
'away': 1187,
'money': 4593,
'worries': 7628,
'per': 5173,
'request': 5747,
'maangalyam': 4319,
'alaipayuthe': 910,
'has': 3371,
'set': 6059,
'callertune': 1629,
'callers': 1628,
'copy': 2021,
'friends': 3017,
'work': 7618,
'yetunde': 7721,
'hasn': 3373,
'sent': 6043,
'her': 3441,
'bother': 1459,
'sending': 6036,
'over': 5054,
'involve': 3757,
'yourself': 7743,
'anything': 1017,
'shouldn': 6150,
'imposed': 3668,
'place': 5256,
'apologise': 1030,
'until': 7206,
'545': 568,
'dinner': 2349,
'together': 6967,
'arrested': 1091,

```
'possession': 5352,
'shit': 6121,
'pm': 5292,
'meanwhile': 4442,
'suite': 6612,
'xavier': 7673,
'decided': 2214,
'give': 3163,
'us': 7239,
'seconds': 5998,
'warning': 7414,
'samantha': 5911,
'was': 7416,
'coming': 1924,
'playing': 5275,
'jay': 3826,
'guitar': 3300,
'impress': 3671,
'doug': 2443,
'realizes': 5636,
'live': 4182,
'anymore': 1012,
'shall': 6085,
'pouch': 5367,
'prof': 5467,
'passed': 5129,
'papers': 5098,
'sem': 6029,
'congrats': 1977,
'student': 6560,
'enna': 2623,
'kalaachutaarama': 3918,
'mrng': 4639,
'come': 1918,
'mahal': 4345,
'stop': 6516,
'decimal': 2216,
'practising': 5386,
'curtsey': 2135,
'philosophy': 5212,
'happen': 3350,
'emigrated': 2597,
'something': 6327,
'maybe': 4429,
'30': 431,
```

```
'bit': 1388,
'hopeful': 3525,
'might': 4499,
'sch': 5960,
'yar': 7699,
'salon': 5909,
'boring': 1452,
'told': 6974,
'already': 939,
'whenevr': 7508,
'sad': 5894,
'gray': 3254,
'im': 3651,
'listn': 4176,
'watevr': 7430,
'wanna': 7401,
'jus': 3903,
'walk': 7386,
'wid': 7533,
'little': 4181,
'while': 7517,
'promise': 5479,
'll': 4190,
'bring': 1524,
'yo': 7728,
'carlos': 1678,
'oops': 4984,
'haf': 3316,
'enuff': 2637,
'check': 1768,
'tell': 6780,
'thanx': 6830,
'schedule': 5961,
'town': 7028,
'any': 1009,
'frm': 3023,
'wats': 7433,
'matter': 4418,
'send': 6034,
'plus': 5289,
'mode': 4577,
'hope': 3523,
'having': 3391,
'semester': 6030,
'wish': 7566,
```

```
'best': 1344,
'greatness': 3260,
'ttyl': 7093,
'omw': 4960,
'castor': 1695,
'fr': 2987,
'ndship': 4745,
'needle': 4761,
'clock': 1858,
'though': 6885,
'same': 5913,
'able': 770,
'evn': 2697,
'itz': 3796,
'4few': 526,
'alwys': 948,
'stay': 6480,
'conected': 1968,
'9t': 753,
'guys': 3305,
'getting': 3143,
'were': 7485,
'staying': 6483,
'mcr': 4433,
'wn': 7594,
'hurt': 3595,
'by': 1603,
'prsn': 5504,
'who': 7521,
'close': 1860,
'fight': 2848,
'wit': 7573,
'dem': 2255,
'somtimes': 6336,
'saves': 5945,
'being': 1326,
'quiet': 5562,
'leaves': 4092,
'nothin': 4863,
'eveb': 2678,
'hand': 3335,
'drama': 2454,
'enough': 2625,
'family': 2784,
'struggling': 6554,
```

'hot': 3541,
'sun': 6623,
'strange': 6534,
'reason': 5642,
'should': 6148,
'ego': 2564,
'actually': 818,
'necessity': 4754,
'wait': 7378,
'serious': 6051,
'reppurcussions': 5744,
'joking': 3873,
'seriously': 6052,
'shock': 6130,
'cos': 2030,
'telling': 6781,
'leona': 4109,
'dat': 2185,
'almost': 935,
'liao': 4122,
'ask': 1111,
'abt': 779,
'reaction': 5625,
'wah': 7373,
'okie': 4949,
'muz': 4691,
'use': 7243,
'unlimited': 7196,
'let': 4116,
'compass': 1938,
'mind': 4511,
'map': 4381,
'soul': 6358,
'guide': 3296,
'never': 4785,
'loose': 4234,
'gnun': 3179,
'via': 7302,
'way2sms': 7439,
'com': 1912,
'number': 4887,
'awarded': 1186,
'ukp': 7149,
'09061790125': 205,
'landline': 4037,

```
'3030': 437,
'150ppm': 314,
'stupid': 6571,
'challenge': 1732,
'god': 3185,
'write': 7646,
'instead': 3731,
'respond': 5765,
'immed': 3659,
'fighting': 2849,
'easy': 2532,
'lose': 4238,
'fightng': 2850,
'some1': 6319,
'dificult': 2335,
'still': 6505,
'lanre': 4045,
'fakeye': 2776,
'eckankar': 2539,
'mail': 4349,
'box': 1468,
'gonna': 3201,
'tacos': 6704,
'open': 4985,
'tuition': 7097,
'empty': 2602,
'slots': 6257,
'driving': 2473,
'x2': 7668,
'bought': 1463,
'test': 6807,
'yesterday': 7718,
'lets': 4117,
'exact': 2708,
'ovulate': 5060,
'2u': 424,
'3wks': 481,
'fret': 3010,
'relax': 5703,
'past': 5135,
'history': 3472,
'restaurant': 5772,
'eng': 2612,
'rocking': 5831,
'ashes': 1105,
```

'shoot': 6134,
'big': 1367,
'loads': 4196,
'ready': 5630,
'tried': 7064,
'reply': 5740,
'video': 7309,
'750': 630,
'mins': 4523,
'camcorder': 1637,
'08000930705': 53,
'ringtone': 5815,
'club': 1870,
'singles': 6204,
'chart': 1753,
'each': 2520,
'top': 7007,
'quality': 5552,
'message': 4484,
'enjoyed': 2619,
'content': 1995,
'61610': 589,
'unsubscribe': 7203,
'08712400602450p': 105,
'provided': 5501,
'tones2you': 6990,
'thinkin': 6870,
'everywhere': 2695,
'dirt': 2358,
'floor': 2916,
'windows': 7551,
'even': 2679,
'shirt': 6119,
'sometimes': 6329,
'mouth': 4627,
'comes': 1920,
'flowing': 2922,
'dream': 2460,
'without': 7578,
'half': 3326,
'chores': 1819,
'joy': 3883,
'lots': 4246,
'things': 6867,
'must': 4685,

```
'exist': 2726,
'hail': 3320,
'mist': 4544,
'done': 2425,
'become': 1302,
'tms': 6954,
'widelive': 7534,
'index': 3698,
'wml': 7593,
'id': 3622,
'820554ad0a1705572711': 666,
'ay': 1195,
'wana': 7400,
'sat': 5932,
'wkg': 7587,
'well': 7478,
'slightly': 6252,
'disastrous': 2367,
'fav': 2809,
'darlings': 2180,
'coffee': 1891,
'wld': 7592,
'tomorrow': 6985,
'always': 947,
'connection': 1980,
'bw': 1600,
'would': 7637,
'wot': 7635,
'drinkin': 2466,
'dancin': 2167,
'cinema': 1828,
'gudnite': 3289,
'tc': 6753,
'practice': 5384,
'forget': 2962,
'owns': 5067,
'property': 5492,
'boy': 1484,
'passionate': 5132,
'kiss': 3981,
'thanks': 6827,
'wonderful': 7607,
'burial': 1579,
'mojibiola': 4584,
'current': 2132,
```

'leading': 4084,
'bid': 1366,
'151': 318,
'pause': 5144,
'auction': 1153,
'care': 1667,
'08718726270': 144,
'anywhere': 1022,
'damn': 2163,
'hunting': 3591,
'eat': 2533,
'old': 4955,
'airport': 899,
'road': 5824,
'630': 594,
'oredi': 5011,
'pple': 5376,
'lunch': 4291,
'alone': 936,
'loving': 4267,
'doesnt': 2406,
'stressful': 6540,
'gr8': 3233,
'ass': 1121,
'enjoy': 2618,
'doggy': 2412,
'applausestore': 1037,
'monthlysubscription': 4604,
'50p': 554,
'msg': 4642,
'max6': 4425,
'csc': 2110,
'web': 7451,
'2stop': 421,
'sleep': 6241,
'hm': 3482,
'headache': 3397,
'anyone': 1013,
'sorry': 6350,
'feel': 2821,
'bad': 1214,
'totally': 7022,
'rejected': 5699,
'kinda': 3974,
'tbs': 6752,

```
'persolvo': 5189,
'chasing': 1756,
'sept': 6047,
'38': 463,
'paying': 5151,
'information': 3709,
'ignore': 3639,
'kath': 3932,
'manchester': 4370,
'death': 2209,
'leave': 4091,
'note': 4861,
'says': 5950,
'robs': 5829,
'fault': 2808,
'avenge': 1176,
'hours': 3549,
'internet': 3743,
'weather': 7450,
'dude': 2498,
'teresa': 6797,
'okay': 4945,
'didnt': 2321,
'these': 6857,
'people': 5172,
'received': 5655,
'package': 5074,
'dec': 2211,
'thot': 6883,
'ld': 4078,
'fantastic': 2792,
'reading': 5629,
'bam': 1231,
'aid': 891,
'usmle': 7251,
'famamus': 2782,
'itself': 3794,
'however': 3558,
'she': 6099,
'stays': 6484,
'ors': 5020,
'stool': 6515,
'crazy': 2076,
'frog': 3031,
'sound': 6359,
```

```
        'poly': 5324,
        'mad1': 4331,
        'real': 5631,
        'mad2': 4332,
        'sounds': 6361,
        'gbp': 3109,
        'syllabus': 6690,
        'aathi': 762,
        'thought': 6886,
        'expert': 2737,
        'creepy': 2087,
        'reality': 5633,
        'fantasy': 2793,
        'show': 6155,
        '08707509020': 91,
        '20p': 362,
        'min': 4508,
        ...}
```

In [58]:
```python
Xtrain_vect=cv.transform(Xtrain.values)
```

In [59]:
```python
Xtrain_vect.toarray()
```

Out[59]:
```
array([[0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0],
       ...,
       [0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0]], dtype=int64)
```

In [60]:
```python
xtest_vect=cv.transform(Xtest)
```

In [61]:
```python
xtest_vect.toarray()
```

Out[61]:
```
array([[0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0],
       ...,
```

```
        [0, 0, 0, ..., 0, 0, 0],
        [0, 0, 0, ..., 0, 0, 0],
        [0, 0, 0, ..., 0, 0, 0]], dtype=int64)
```

In [62]:
```python
from sklearn.naive_bayes import MultinomialNB
```

In [63]:
```python
mmb=MultinomialNB()
```

In [64]:
```python
mmb.fit(Xtrain_vect,ytrain)
```

Out[64]: MultinomialNB()

In [65]:
```python
mmb.score(xtest_vect,ytest)*100
```

Out[65]: 98.83408071748879

In [68]:
```python
a=["hii how are you  my name is teja"]
```

In [69]:
```python
msg=cv.transform(a)
```

In [70]:
```python
msg.toarray()
```

Out[70]: array([[0, 0, 0, ..., 0, 0, 0]], dtype=int64)

In [71]:
```python
mmb.predict(msg)
```

Out[71]: array([0], dtype=int64)

In [80]:
```python
b=["Congrats! 1 year special cinema pass for 2 is yours. call 09061209465 now! C Suprman V, Matrix3, StarWars3, etc a
```

```
In [81]:   msg1=cv.transform(b)
```

```
In [82]:   msg.toarray()
```

Out[82]: array([[0, 0, 0, ..., 0, 0, 0]], dtype=int64)

```
In [83]:   mmb.predict(msg1)
```

Out[83]: array([1], dtype=int64)

```
In [ ]:
```