In [18]:
```
import numpy as np
import pandas as pd
import gensim
```

In [2]:
```
df=pd.read_csv('C:\\Users\\LENOVO\\Documents\\PHONE_REVIEW.CSV')
```

In [39]:
```
df.head(5)
```

Out[39]:

	asin	name	rating	date	verified	title	body	helpfulVotes
0	B0000SX2UC	Janet	3	October 11, 2005	False	Def not best, but not worst	I had the Samsung A600 for awhile which is abs...	1.0
1	B0000SX2UC	Luke Wyatt	1	January 7, 2004	False	Text Messaging Doesn't Work	Due to a software issue between Nokia and Spri...	17.0
2	B0000SX2UC	Brooke	5	December 30, 2003	False	Love This Phone	This is a great, reliable phone. I also purcha...	5.0
3	B0000SX2UC	amy m. teague	3	March 18, 2004	False	Love the Phone, BUT...!	I love the phone and all, because I really did...	1.0
4	B0000SX2UC	tristazbimmer	4	August 28, 2005	False	Great phone service and options, lousy case!	The phone has been great for every purpose it ...	1.0

In [4]:
```
df.isnull()
```

Out[4]:

	asin	name	rating	date	verified	title	body	helpfulVotes
0	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...
67981	False	False	False	False	False	False	False	False
67982	False	False	False	False	False	False	False	False
67983	False	False	False	False	False	False	False	True
67984	False	False	False	False	False	False	False	True
67985	False	False	False	False	False	False	False	True

67986 rows × 8 columns

In [6]:
```
df.isnull().sum()
```

Out[6]:
```
asin                0
name                2
rating              0
date                0
verified            0
title              14
body               21
helpfulVotes    40771
dtype: int64
```

In [9]:
```
df.shape
```

Out[9]:
```
(67986, 8)
```

In [10]:
```
df.dtypes
```

Out[10]:
```
asin             object
name             object
rating            int64
date             object
verified           bool
title            object
body             object
helpfulVotes    float64
dtype: object
```

In [12]:
```
df1=df.dropna()
```

In [13]:
```
df1.shape
```

Out[13]:
```
(27206, 8)
```

In [14]:
```
df1.head(10)
```

Out[14]:

	asin	name	rating	date	verified	title	body	helpfulVotes
0	B0000SX2UC	Janet	3	October 11, 2005	False	Def not best, but not worst	I had the Samsung A600 for awhile which is abs...	1.0
1	B0000SX2UC	Luke Wyatt	1	January 7, 2004	False	Text Messaging Doesn't Work	Due to a software issue between Nokia and Spri...	17.0
2	B0000SX2UC	Brooke	5	December 30, 2003	False	Love This Phone	This is a great, reliable phone. I also purcha...	5.0
3	B0000SX2UC	amy m. teague	3	March 18, 2004	False	Love the Phone, BUT...!	I love the phone and all, because I really did...	1.0
4	B0000SX2UC	tristazbimmer	4	August 28, 2005	False	Great phone service and options, lousy case!	The phone has been great for every purpose it ...	1.0
6	B0000SX2UC	the cell phone store owner	5	April 16, 2004	False	Wanna cool Nokia? You have it here!	Cool. Cheap. Color: 3 words that describe the ...	2.0
7	B0000SX2UC	Matt	4	April 3, 2004	False	Problem with 3588i universal headset	The 3599i is overall a nice phone, except that...	2.0
8	B0000SX2UC	Charles Cook	5	November 24, 2003	False	cool phone!!!!!!!!	I've never owned a Nokia phone before, so this...	7.0
9	B0000SX2UC	Amazon Customer	3	February 2, 2004	False	Pissed off-a little bit	ok well im in school and i need the text messa...	3.0
10	B0000SX2UC	habblie	4	December 25, 2004	False	works great, but don't dropt it	I've had this phone for over a year and I real...	1.0

In [17]:
```
df1.title[10]
```

Out[17]:
```
"works great, but don't dropt it"
```

In [19]:
```
new_df1=df1.title.apply(gensim.utils.simple_preprocess)
```

In [21]:
```
new_df1
```

Out[21]:
```
0                        [def, not, best, but, not, worst]
1                           [text, messaging, doesn, work]
2                                     [love, this, phone]
3                                  [love, the, phone, but]
4        [great, phone, service, and, options, lousy, c...
                               ...
67978                         [candy, bar, phone, is, back]
67979                                     [updated, review]
67980         [from, iphone, to, android, and, loving, it]
67981    [awesome, phone, but, finger, scanner, is, big...
67982                                     [simply, amazing]
Name: title, Length: 27206, dtype: object
```

In [22]:
```
model = gensim.models.Word2Vec(
    window=10,
    min_count=4,
    workers=8)
```

In [25]:
```
model.build_vocab(new_df1,progress_per=1000)
```

In [26]:
```
model.epochs
```

Out[26]:
```
5
```

In [27]:
```
model.corpus_count
```

Out[27]:
```
27206
```

In [28]:
```
model.train(new_df1,total_examples=model.corpus_count,epochs=model.epochs)
```

Out[28]:
```
(443676, 672315)
```

In [29]:
```
model.save("./mobile_phone_review")
```

In [30]:
```
#to experiment our model
model.wv.most_similar('good')
```

Out[30]:
```
[('nice', 0.966987669467926),
 ('perfect', 0.9508532881736755),
 ('well', 0.9398593306541443),
 ('great', 0.939740002155304),
 ('awesome', 0.935762345790863),
 ('pleased', 0.9262329936027527),
 ('shape', 0.9232280850410461),
 ('reasonable', 0.9226639270782471),
 ('excellent', 0.9220534563064575),
 ('amazing', 0.9219745397567749)]
```

In [32]:
```
model.wv.most_similar('bad')
```

Out[32]:
```
[('terrible', 0.9721834063529968),
 ('long', 0.9673416018486023),
 ('short', 0.9634677767753601),
 ('dies', 0.9624043703079224),
 ('customer', 0.9607540965080261),
 ('sucks', 0.9578807353973389),
 ('horrible', 0.9563532471656799),
 ('replaceable', 0.952724814414978),
 ('reader', 0.9523420333862305),
 ('though', 0.95054030418396)]
```

In [42]:
```
model.wv.most_similar('love')
```

Out[42]:
```
[('loved', 0.8968225121498108),
 ('used', 0.8955539464950562),
 ('absolutely', 0.8873627781867981),
 ('is', 0.8842310905456543),
 ('hate', 0.8818738460540771),
 ('my', 0.8813616633415222),
 ('its', 0.8795970678329468),
 ('favorite', 0.8770487308502197),
 ('refurbished', 0.8766623735427856),
 ('like', 0.8761147856712341)]
```

In [31]:
```
#similarity between two words
model.wv.similarity(w1="cheap",w2="good")
```

Out[31]:
```
0.88509977
```

In [33]:
```
model.wv.similarity(w1="cheap",w2="samsung")
```

Out[33]:
```
0.7861464
```

In [ ]:
```

```