

## Information about data:

->We have the amazon reviews dataset from kaggle

->Reviews are given for the product

->The features of the data were:

Id

ProductId- unique identifier for the product

UserId- unique identifier for the user

ProfileName

HelpfulnessNumerator- number of users who found the review helpful

HelpfulnessDenominator- number of users who indicated whether they found the review helpful or not

Score-rating between 1 and 5

Time-timestamp for the review

Summary- brief summary of the review

Text- text of the review

-> Based on the score of the review we classify them into positive and negative

Number of reviews: 568,454



## Objective:

-> Clean the data and to get the complete vocabulary

-> To get frequencies of each word in the vocabulary

-> Convert the text into the values of the word frequencies

-> To try different layers of LSTM, with different number of LSTM units in each layer

## Loading the required libraries to process

In [60]:

```
import sqlite3
import nltk
import string
from sklearn.metrics import accuracy_score
from nltk.stem.porter import PorterStemmer
```

## Loading the data

In [2]:

```
con = sqlite3.connect("database.sqlite")
data = pd.read_sql_query("SELECT * FROM Reviews", con)
```

## Analysis about the given data:

- >The shape of data
- >Number of dimension of the data
- >Number of attributes and their names
- >Sample subset of data

In [96]:

```
print(data.shape)
print(data.ndim)
print(data.columns)
print(data.head(5))
```

(568454, 10)

2

```
Index(['Id', 'ProductId', 'UserId', 'ProfileName', 'HelpfulnessNumerator',
       'HelpfulnessDenominator', 'Score', 'Time', 'Summary', 'Text'],
      dtype='object')
```

		Id	ProductId	UserId	ProfileName	\
0	1	B001E4KFG0	A3SGXH7AUHU8GW		delmartian	
1	2	B00813GRG4	A1D87F6ZCVE5NK		dll pa	
2	3	B000LQOCH0	ABXLMWJIXXAIN	Natalia Corres	"Natalia Corres"	
3	4	B000UA0QIQ	A395BORC6FGVXV		Karl	
4	5	B006K2ZZ7K	A1UQRSCLF8GW1T	Michael D. Bigham	"M. Wassir"	

		HelpfulnessNumerator	HelpfulnessDenominator	Score	Time	\
0		1	1	positive	1303862400	
1		0	0	negative	1346976000	
2		1	1	positive	1219017600	
3		3	3	negative	1307923200	
4		0	0	positive	1350777600	

	Summary	Text
0	Good Quality Dog Food	I have bought several of the Vitality canned d...
1	Not as Advertised	Product arrived labeled as Jumbo Salted Peanut...
2	"Delight" says it all	This is a confection that has been around a fe...
3	Cough Medicine	If you are looking for the secret ingredient i...
4	Great taffy	Great taffy at a great price. There was a wid...

#### NOTE:

-> We have the reviews in the form of text and ratings are in the range of 1 to 5

-> Converting the rating score into positive, negative based on their score

-> Considering the reviews with rating score greater than 3 as positive and less than or equal to three as negative

In [97]:

```
def reviews(x):
    if x>3:
        return 'positive'
    return 'negative'
```

In [99]:

```
data.head(5)
```

Out[99]:

	Id	ProductId	UserId	ProfileName	HelpfulnessNumerator	HelpfulnessDenominator
0	1	B001E4KFG0	A3SGXH7AUHU8GW	delmartian	1	1
1	2	B00813GRG4	A1D87F6ZCVE5NK	dll pa	0	0
2	3	B000000000	A1D87F6ZCVE5NK	Natalia Corres	1	1

2	3	B000LQOCH0	ABXLMWJIXXAIN	ProfileName	1	1
	Id	ProductId	UserId	Corres	HelpfulnessNumerator	HelpfulnessDenominator
3	4	B000UA0QIQ	A395BORC6FGVXV	Karl	3	3
4	5	B006K2ZZ7K	A1UQRSCLF8GW1T	Michael D. Bigham "M. Wassir"	0	0

Exploratory data analysis

Deduplication: Removing duplicates

In [100]:

```
sorteddata = data.sort_values('ProductId',axis=0,inplace=False,ascending=True,kind='quicksort',na_position='last')
```

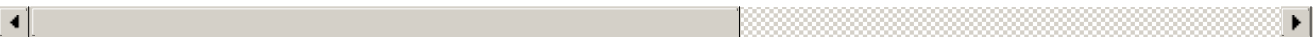
In [101]:

```
sorteddata.head(6)
```

Out[101]:

	Id	ProductId	UserId	ProfileName	HelpfulnessNumerator	HelpfulnessDenominator
150528	150529	0006641040	A25ACLV5KPB4W	Matt Hetling "Matt"	0	1
150506	150507	0006641040	A1S4A3IQ2MU7V4	sally sue "sally sue"	1	1
150505	150506	0006641040	A2IW4PEEKO2R0U	Tracy	1	1
				Jason A		

	Id	ProductId	UserId	ProfileName	HelpfulnessNumerator	Helpful
<b>150504</b>	150505	0006641040	A2PTSM496CF40Z	"Nobody made a greater mistak...	1	1
<b>150503</b>	150504	0006641040	AQEYF1AXARWJZ	Les Sinclair "book maven"	1	1
<b>150502</b>	150503	0006641040	A3R5XMPFU8YZ4D	Her Royal Motherliness "Nana"	1	1



In [102]:

```
final =
sorteddata.drop_duplicates(subset=("UserId", 'Time', 'Text', 'ProfileName'), ke
ep='first', inplace=False)
```

In [103]:

```
final.shape
```

Out[103]:

```
(393933, 10)
```

In [104]:

```
data.shape
```

Out[104]:

```
(568454, 10)
```

In [105]:

```
(final['Id'].size/data['Id'].size)*100
```

Out[105]:

```
69.29901100176971
```

In [106]:

```
final.shape
```

Out[106]:

```
(393933, 10)
```

Taking subset of data from the dataset

In [107]:

```
final = final.sample(n=100000)
final.shape
```

Out[107]:

```
(100000, 10)
```

Keeping only Reviews, Score and discarding the remaining features

In [108]:

```
final = final[['Text', 'Score']]
print(final.shape)
final.head(3)
```

```
(100000, 2)
```

Out[108]:

	Text	Score
460022	(Mr. Master Reviewer ^.^)  This Ener...	positive
544096	I guess I'm not lucky the other guys.in my cas...	negative
564627	We started purchasing Eight O'clock Coffee abo...	positive

Converting all text to lower case

In [109]:

```
final['Text'] = final['Text'].str.lower()
```

In [110]:

```
nu_top_words = 5000
```

In [111]:

```
sentences = final['Text'].str.cat(sep = ' ')
```

In [112]:

```
words = nltk.tokenize.word_tokenize(sentences)
```

In [113]:

```
nltk.download('punkt')
```

```
[nltk_data] Downloading package punkt to C:\Users\Anil
[nltk_data] Chowdary\AppData\Roaming\nltk_data...
[nltk_data] Package punkt is already up-to-date!
```

Out[113]:

```
True
```

In [114]:

```
len(words)
```

Out[114]:

9542183

In [115]:

```
words
```

Out[115]:

```
['(',  
'mr.',  
'master',  
'reviewer',  
'^.^',  
)',  
'<',  
'br',  
'/',  
'>',  
'<',  
'br',  
'/',  
'>',  
'this',  
'energy',  
'gel',  
'24',  
'pack',  
'is',  
'simply',  
'the',  
'best',  
'buy',  
'for',  
'any',  
'person',  
'who',  
'is',  
'in',  
'doing',  
'marathons',  
'or',  
'simply',  
'training',  
'the',  
'body',  
'.',  
'i',  
'have',  
'gone',  
'through',  
'at',  
'least',  
'8',  
'orders',  
'of',  
'this',
```

'stuff',  
'until',  
'i',  
'finally',  
'realized',  
'how',  
'much',  
'of',  
'a',  
'killer',  
'it',  
'can',  
'be',  
'on',  
'the',  
'average',  
'user',  
"s",  
'wallet',  
'.',  
'if',  
'these',  
'things',  
'were',  
'cheaper',  
'or',  
'if',  
'i',  
'could',  
'get',  
'them',  
'for',  
'an',  
'affordable',  
'price',  
'',  
'they',  
'would',  
'trully',  
'be',  
'amazing',  
'but',  
'either',  
'way',  
'',  
'this',  
'stuff',  
'helps',  
'!',  
'flavors',  
'and',  
'extremely',  
'tasty',  
'and',  
'regardless',  
'of',  
'the',  
'type',  
'',  
'they',  
...



'all',  
'go',  
'down',  
'and',  
'about.',  
'<',  
'br',  
'/',  
'>',  
'<',  
'br',  
'/',  
'>',  
'my',  
'favorite',  
'beyond',  
'all',  
'the',  
'flavors',  
'is',  
'`',  
'espresso',  
'love',  
''',  
'because',  
'it',  
'has',  
'the',  
'most',  
'caffeine',  
'in',  
'one',  
'gel',  
'shot',  
'and',  
'is',  
'a',  
'killer',  
'(',  
'in',  
'a',  
'good',  
'way',  
'')',  
'for',  
'coffee',  
'lovers',  
'!',  
'if',  
'interested',  
'in',  
'discovering',  
'which',  
'flavor',  
'is',  
'best',  
'',  
'give',  
'the',  
'`',  
'mixed'

'mixed',  
''',  
'combo',  
'a',  
'shot',  
'.',  
'i',  
'have',  
'uploaded',  
'pictures',  
'of',  
'all',  
'flavors',  
'you',  
'get',  
'in',  
'this',  
'pack',  
'.',  
'although',  
'\$',  
'6',  
'more',  
'expensive',  
'than',  
'a',  
'specific',  
'flavor',  
'buy',  
',',  
'it',  
'gives',  
'you',  
'more',  
'variety',  
'^.^',  
'i',  
'guess',  
'i',  
"m",  
'not',  
'lucky',  
'the',  
'other',  
'guys.in',  
'my',  
'case',  
'as',  
'soon',  
'i',  
'opened',  
'the',  
'bag',  
'i',  
'knew',  
'right',  
'way',  
'i',  
'have',  
'a',  
'problem beans'

problem.beans ,  
'smell',  
'like',  
'dirt/soil',  
,',  
'plus',  
'no',  
'coffee',  
'taste',  
'at',  
'all.i',  
'tried',  
'many',  
'brands',  
'before',  
,',  
'but',  
'this',  
'one',  
'by',  
'far',  
'is',  
'the',  
'worst.good',  
'thing',  
'i',  
'still',  
'have',  
'a',  
'bag',  
'of',  
'weaver',  
"s",  
'coffee',  
'we',  
'started',  
'purchasing',  
'eight',  
"o'clock",  
'coffee',  
'about',  
'ten',  
'years',  
'ago',  
'when',  
'we',  
'saw',  
'it',  
'in',  
'our',  
'local',  
'grocery',  
'story',  
,',  
'we',  
'like',  
'the',  
'whole',  
'beans',  
'and',  
'also'.

also ,  
'the',  
'decaf',  
'.',  
'my',  
'husband',  
'and',  
'i',  
'have',  
'decided',  
'that',  
'it',  
'is',  
'probably',  
'the',  
'most',  
'under-rated',  
'coffee',  
'ever',  
'.',  
'it',  
'is',  
'rich',  
',',  
'flavorful',  
',',  
'and',  
'exciting',  
'.',  
'the',  
'flavor',  
'is',  
'distinct',  
'and',  
'original',  
'.',  
'it',  
'is',  
'not',  
'starbucks',  
',',  
'but',  
'it',  
'is',  
'far',  
'better',  
'and',  
'much',  
'less',  
'expensive',  
'.',  
'we',  
'have',  
'found',  
'eight',  
"o'clock",  
'coffee',  
'to',  
'be',  
'consistantly',  
'fresh',

','  
'and',  
'the',  
'aroma',  
'while',  
'is',  
'is',  
'brewing',  
','  
'literally',  
'defines',  
'comfort',  
'and',  
'contentment',  
'.'  
'i',  
'can',  
'run',  
'around',  
'with',  
'a',  
'steaming',  
'cup',  
'in',  
'the',  
'morning',  
'and',  
'feel',  
'confident',  
'that',  
'my',  
'day',  
'is',  
'going',  
'to',  
'be',  
'just',  
'perfect',  
'.'  
'i',  
'conclude',  
'by',  
'saying',  
'that',  
'this',  
'coffee',  
'will',  
'steal',  
'you',  
'away',  
'from',  
'other',  
'much',  
'more',  
'expensive',  
'brands',  
'and',  
'you',  
'will',  
'be',

'an',  
'eight',  
"o'clock",  
'convert',  
'for',  
'life',  
'.',  
'these',  
'are',  
'awesome',  
'and',  
'they',  
'do',  
"n't",  
'have',  
'the',  
'usual',  
'high',  
'sodium',  
'(',  
'salt',  
)',  
'content',  
'that',  
'many',  
'snacks',  
'have',  
'.',  
'outstanding',  
'taste',  
'and',  
'very',  
'healthy',  
'.',  
'great',  
'product',  
'',  
'ca',  
"n't",  
'believe',  
'this',  
'has',  
"n't",  
'been',  
'available',  
'before.',  
'<',  
'br',  
'/',  
'>',  
'<',  
'br',  
'/',  
'>',  
'steve',  
'antos',  
'medaglia',  
"d'oro",  
'has',  
'to',

'be',  
'the',  
'worst',  
'coffee',  
'i',  
"ve",  
'ever',  
'had',  
'.',  
'it',  
'had',  
'an',  
'old',  
'musty',  
'flavor',  
'and',  
'odor',  
'.',  
'you',  
'get',  
'what',  
'pay',  
'for',  
'.',  
'strong',  
'and',  
'lovely',  
'lemon',  
'flavor',  
'in',  
'a',  
'crispy',  
'little',  
'square',  
'.',  
'they',  
'are',  
'delicious',  
'and',  
'tangy',  
'.',  
'i',  
'keep',  
'them',  
'in',  
'the',  
'refrigerator',  
'because',  
'i',  
'think',  
'they',  
'taste',  
'better',  
'cold',  
'.',  
'just',  
'grab',  
'a',  
'few',  
'of',

'these',  
'little',  
'treats',  
'to',  
'satisfy',  
'your',  
'cookie',  
'cravings',  
'.',  
'yum',  
'.',  
'most',  
'individuals',  
'might',  
'compare',  
'these',  
'candies',  
'to',  
'cough',  
'drops',  
'',  
'heavy',  
'on',  
'the',  
'ginger',  
'and',  
'lite',  
'on',  
'mint',  
'.',  
'if',  
'you',  
'enjoy',  
'ginger',  
'',  
'these',  
'hard',  
'candies',  
'will',  
'satisfy',  
'.',  
'i',  
'was',  
'not',  
'sure',  
'this',  
'product',  
'would',  
'live',  
'up',  
'to',  
'the',  
'ads',  
'',  
'but',  
'it',  
'gave',  
'me',  
'the',  
'energy',  
..



'boost',  
'i',  
'needed',  
'when',  
'i',  
'needed',  
'it',  
'when',  
'i',  
'had',  
'to',  
'work',  
'the',  
'overnight',  
'shift',  
'.',  
'what',  
'a',  
'joy',  
'to',  
'have',  
'iced',  
'tea',  
'in',  
'just',  
'a',  
'couple',  
'minutes',  
'.',  
'the',  
'flavors',  
'are',  
'wonderful',  
'and',  
'so',  
'good',  
'in',  
'this',  
'hot',  
'weather.',  
'<',  
'br',  
'/',  
'>',  
'i',  
'recommend',  
'this',  
'sampler',  
'to',  
'anyway',  
'that',  
'loves',  
'tea',  
'.',  
'my',  
'favorite',  
'sugarless',  
'candy',  
'',  
'although',  
'',

'1',  
'find',  
'that',  
'most',  
'stores',  
'do',  
'not',  
'carry',  
'them',  
'or',  
'they',  
'are',  
'always',  
'out',  
'of',  
'stock',  
'.',  
'this',  
'is',  
'why',  
'i',  
'now',  
'order',  
'them',  
'direct',  
'through',  
'amazon.com',  
'and',  
'receive',  
'regular',  
'shipments.',  
'<',  
'br',  
'/',  
'>',  
'<',  
'br',  
'/',  
'>',  
'valerie',  
'',  
'valley',  
'springs',  
'',  
'ca',  
'there',  
'are',  
'two',  
'kinds',  
'of',  
'tea',  
':',  
'twinings',  
'',  
'and',  
'all',  
'others',  
'.',  
'twinings',  
'english',  
'breakfast'

breakfast ,  
'tea',  
'has',  
'a',  
'welcoming',  
'aroma',  
'and',  
'flavor',  
'that',  
'sets',  
'it',  
'apart',  
'from',  
'wannabee',  
'competitors.',  
'<',  
'br',  
'/',  
'>',  
'<',  
'br',  
'/',  
'>',  
'amazon',  
"s",  
'price',  
'is',  
'amazing',  
,',  
'which',  
'worked',  
'out',  
'to',  
'slightly',  
'less',  
'than',  
'11',  
'cents',  
'per',  
'tea',  
'bag',  
'.',  
'the',  
'next',  
'best',  
'price',  
'i',  
"ve",  
'seen',  
'recently',  
'was',  
'closer',  
'to',  
'15',  
'cents',  
'at',  
'a',  
'dod',  
'commissary',  
,',  
'with'.

with ,  
'la-area',  
'stores',  
'routinely',  
'in',  
'the',  
'19-20',  
'cent/bag',  
'range',  
'.',  
'my',  
'order',  
'was',  
'nicely',  
'packaged',  
'and',  
'arrived',  
'within',  
'2',  
'days',  
'of',  
'ordering',  
'.',  
'this',  
'is',  
'extreme',  
'value',  
'.',  
'ghost',  
'peppers',  
'have',  
'a',  
'great',  
'flavor',  
'',  
'i',  
'think',  
'this',  
'sauces',  
'captured',  
'that',  
'without',  
'blowing',  
'it',  
'away',  
'with',  
'heat',  
'.',  
'do',  
'n't',  
'get',  
'me',  
'wrong',  
'',  
'it',  
'has',  
'decent',  
'heat',  
'but',  
'its',  
'not',

even',  
'close',  
'to',  
'the',  
'pain',  
'level',  
'of',  
'the',  
'insanity',  
'sauce',  
'.',  
'this',  
'is',  
'a',  
'nice',  
'step',  
'up',  
'from',  
'something',  
'like',  
'tapitio',  
';',  
'its',  
'great',  
'on',  
'chicken',  
',',  
'burritos',  
',',  
'chili',  
',',  
'etc',  
'.',  
'if',  
'you',  
'want',  
'to',  
'add',  
'some',  
'great',  
'ghost',  
'pepper',  
'flavor',  
'and',  
'a',  
'bit',  
'of',  
'heat',  
'to',  
'your',  
'food',  
'this',  
'is',  
'the',  
'stuff',  
'you',  
'want',  
'.',  
'this',  
'drink',

'is',  
'far',  
'too',  
'sweet',  
'.',  
'i',  
'ended',  
'up',  
'watering',  
'it',  
'down',  
'to',  
'drink',  
'it',  
'.',  
'my',  
'husband',  
'never',  
'cared',  
'much',  
'for',  
'my',  
'homemade',  
'biscuits',  
'.',  
'i',  
'did',  
'n't',  
'make',  
'them',  
'often',  
'enough',  
'to',  
'really',  
'get',  
'it',  
'right',  
'like',  
'my',  
'mother',  
'and',  
'grandmother',  
'used',  
'to',  
'do',  
'.',  
'then',  
'i',  
'tried',  
'the',  
'betty',  
'crocker',  
'bisquick',  
'mix',  
'and',  
'wow',  
'what',  
'a',  
'difference',  
'it',

```
'made',
'.',
'we',
'like',
'the',
'rolled',
'method',
'better',
'than',
'the',
'drop',
',',
'but',
'either',
'way',
'these',
'are',
'easy',
',',
'quick',
'and',
'consistently',
'great',
'biscuits',
'.',
'as',
'good',
'a',
'true',
'southern',
'biscuit',
'as',
'even',
'my',
'grandmother',
'used',
'to',
'make',
'!',
'my',
'husband',
'is',
'now',
'a',
'very',
'happy',
'man',
'when',
'i',
'make',
...]
```

In [116]:

```
word_dict = nltk.FreqDist(words)
```

In [117]:

```
word_dict
```

Out[117]:

Out[11]:

```
FreqDist({'(': 35380,  
         'mr.': 114,  
         'master': 85,  
         'reviewer': 905,  
         '^.': 3,  
         ')': 38078,  
         '<': 119401,  
         'br': 113681,  
         '/': 114085,  
         '>': 119781,  
         'this': 120930,  
         'energy': 2436,  
         'gel': 191,  
         '24': 965,  
         'pack': 5145,  
         'is': 132579,  
         'simply': 1582,  
         'the': 327317,  
         'best': 13324,  
         'buy': 13517,  
         'for': 95763,  
         'any': 11399,  
         'person': 1005,  
         'who': 7991,  
         'in': 97932,  
         'doing': 1184,  
         'marathons': 9,  
         'or': 31696,  
         'training': 535,  
         'body': 1285,  
         '.': 420578,  
         'i': 295160,  
         'have': 61096,  
         'gone': 1408,  
         'through': 4418,  
         'at': 31686,  
         'least': 2900,  
         '8': 1784,  
         'orders': 285,  
         'of': 138804,  
         'stuff': 6603,  
         'until': 3095,  
         'finally': 1464,  
         'realized': 446,  
         'how': 6981,  
         'much': 16133,  
         'a': 212402,  
         'killer': 68,  
         'it': 191898,  
         'can': 26492,  
         'be': 33507,  
         'on': 47719,  
         'average': 630,  
         'user': 140,  
         "'s": 52130,  
         'wallet': 54,  
         'if': 29521,  
         'these': 37720,  
         'things': 3336,
```



'were': 14856,  
'cheaper': 2526,  
'could': 10127,  
'get': 18472,  
'them': 34334,  
'an': 15936,  
'affordable': 415,  
'price': 13305,  
' ': 304782,  
'they': 52252,  
'would': 23393,  
'trully': 5,  
'amazing': 2274,  
'but': 64539,  
'either': 2668,  
'way': 8638,  
'helps': 1332,  
'!': 75535,  
'flavors': 6290,  
'and': 224509,  
'extremely': 1333,  
'tasty': 4602,  
'regardless': 205,  
'type': 1819,  
'all': 26394,  
'go': 7429,  
'down': 4623,  
'about.': 37,  
'my': 77680,  
'favorite': 6515,  
'beyond': 438,  
'``': 19215,  
'espresso': 1520,  
'love': 21374,  
' '' ': 24907,  
'because': 14421,  
'has': 22914,  
'most': 7572,  
'caffeine': 1581,  
'one': 29350,  
'shot': 589,  
'good': 34339,  
'coffee': 24726,  
'lovers': 371,  
'interested': 451,  
'discovering': 74,  
'which': 13878,  
'flavor': 24795,  
'give': 7057,  
'mixed': 2041,  
'combo': 230,  
'uploaded': 30,  
'pictures': 111,  
'you': 65155,  
'although': 2563,  
'\$': 7653,  
'6': 3311,  
'more': 23119,  
'expensive': 3672,  
'than': 20356,

'specific': 240,  
'gives': 1393,  
'variety': 2638,  
'guess': 1759,  
'm': 12974,  
'not': 60013,  
'lucky': 307,  
'other': 18784,  
'guys.in': 1,  
'case': 3171,  
'as': 49018,  
'soon': 1309,  
'opened': 1813,  
'bag': 8980,  
'knew': 793,  
'right': 6048,  
'problem.beans': 1,  
'smell': 3418,  
'like': 43441,  
'dirt/soil': 1,  
'plus': 2440,  
'no': 18567,  
'taste': 29782,  
'all.i': 4,  
'tried': 12722,  
'many': 7303,  
'brands': 3482,  
'before': 6858,  
'by': 12570,  
'far': 4790,  
'worst.good': 1,  
'thing': 4772,  
'still': 7787,  
'weaver': 27,  
'we': 24717,  
'started': 2467,  
'purchasing': 1152,  
'eight': 364,  
'o'clock': 136,  
'about': 17449,  
'ten': 419,  
'years': 6316,  
'ago': 2465,  
'when': 23923,  
'saw': 1541,  
'our': 9376,  
'local': 5083,  
'grocery': 3948,  
'story': 254,  
'whole': 5455,  
'beans': 3214,  
'also': 15014,  
'decaf': 1437,  
'husband': 2862,  
'decided': 2606,  
'that': 77976,  
'probably': 3678,  
'under-rated': 1,  
'ever': 5845,  
'rich': 2363,

'flavorful': 1521,  
'exciting': 99,  
'distinct': 279,  
'original': 1671,  
'starbucks': 1671,  
'better': 12252,  
'less': 5944,  
'found': 9480,  
'to': 177086,  
'consistantly': 24,  
'fresh': 5666,  
'aroma': 1535,  
'while': 6262,  
'brewing': 781,  
'literally': 483,  
'defines': 8,  
'comfort': 181,  
'contentment': 2,  
'run': 1520,  
'around': 3932,  
'with': 62737,  
'steaming': 90,  
'cup': 7835,  
'morning': 3205,  
'feel': 3880,  
'confident': 111,  
'day': 7492,  
'going': 4534,  
'just': 30148,  
'perfect': 5677,  
'conclude': 30,  
'saying': 626,  
'will': 22104,  
'steal': 105,  
'away': 3356,  
'from': 26033,  
'convert': 70,  
'life': 1776,  
'are': 54258,  
'awesome': 1798,  
'do': 29096,  
'n't': 51722,  
'usual': 802,  
'high': 4677,  
'sodium': 1700,  
'salt': 4451,  
'content': 1161,  
'snacks': 1592,  
'outstanding': 431,  
'very': 29644,  
'healthy': 5555,  
'great': 28589,  
'product': 27375,  
'ca': 6860,  
'believe': 1904,  
'been': 13235,  
'available': 3174,  
'before.': 48,  
'steve': 10,  
'antos': 1,  
- . . . -

'medaglia': 7,  
"d'oro": 13,  
'worst': 662,  
"ve": 14894,  
'had': 22491,  
'old': 4935,  
'musty': 46,  
'odor': 382,  
'what': 14355,  
'pay': 1635,  
'strong': 4884,  
'lovely': 547,  
'lemon': 1766,  
'crispy': 621,  
'little': 14635,  
'square': 254,  
'delicious': 7522,  
'tangy': 357,  
'keep': 5585,  
'refrigerator': 389,  
'think': 8416,  
'cold': 2095,  
'grab': 397,  
'few': 6454,  
'treats': 4248,  
'satisfy': 456,  
'your': 15222,  
'cookie': 2219,  
'cravings': 260,  
'yum': 836,  
'individuals': 48,  
'might': 3338,  
'compare': 718,  
'candies': 827,  
'cough': 175,  
'drops': 531,  
'heavy': 912,  
'ginger': 2180,  
'lite': 140,  
'mint': 1038,  
'enjoy': 5054,  
'hard': 5179,  
'was': 58367,  
'sure': 5464,  
'live': 1541,  
'up': 16923,  
'ads': 35,  
'gave': 2648,  
'me': 21425,  
'boost': 559,  
'needed': 1461,  
'work': 4654,  
'overnight': 272,  
'shift': 48,  
'joy': 182,  
'iced': 1390,  
'tea': 24078,  
'couple': 2599,  
'minutes': 2748,  
'wonderful': 4394,  
'...': 43310

'so': 43810,  
'hot': 6836,  
'weather.': 12,  
'recommend': 7354,  
'sampler': 354,  
'anyway': 1223,  
'loves': 5302,  
'sugarless': 60,  
'candy': 4002,  
'find': 13304,  
'stores': 4044,  
'carry': 1298,  
'always': 6201,  
'out': 19687,  
'stock': 1376,  
'why': 3419,  
'now': 11531,  
'order': 7781,  
'direct': 282,  
'amazon.com': 928,  
'receive': 620,  
'regular': 4948,  
'shipments.': 3,  
'valerie': 4,  
'valley': 247,  
'springs': 27,  
'there': 15727,  
'two': 8162,  
'kinds': 669,  
' ': 15020,  
'twinings': 327,  
'others': 2718,  
'english': 612,  
'breakfast': 2617,  
'welcoming': 7,  
'sets': 143,  
'apart': 532,  
'wannabee': 3,  
'competitors.': 6,  
'amazon': 14735,  
'worked': 1087,  
'slightly': 1893,  
'11': 484,  
'cents': 473,  
'per': 4069,  
'next': 2718,  
'seen': 1141,  
'recently': 1646,  
'closer': 205,  
'15': 1083,  
'dod': 1,  
'commissary': 31,  
'la-area': 1,  
'routinely': 38,  
'19-20': 1,  
'cent/bag': 1,  
'range': 321,  
'nicely': 771,  
'packaged': 1477,  
'arrived': 3262,  
'within': 1298

within': 1290,  
'2': 7314,  
'days': 3283,  
'ordering': 2305,  
'extreme': 177,  
'value': 1884,  
'ghost': 108,  
'peppers': 490,  
'sauces': 6,  
'captured': 18,  
'without': 6340,  
'blowing': 39,  
'heat': 1837,  
'wrong': 1593,  
'decent': 1107,  
'its': 6238,  
'even': 12969,  
'close': 1528,  
'pain': 454,  
'level': 717,  
'insanity': 78,  
'sauce': 5720,  
'nice': 7433,  
'step': 286,  
'something': 6575,  
'tapatio': 1,  
';': 6452,  
'chicken': 4383,  
'burritos': 57,  
'chili': 1171,  
'etc': 1955,  
'want': 6902,  
'add': 7113,  
'some': 18313,  
'pepper': 1640,  
'bit': 7538,  
'food': 16735,  
'drink': 8983,  
'too': 15819,  
'sweet': 9016,  
'ended': 809,  
'watering': 98,  
'never': 6767,  
'cared': 58,  
'homemade': 923,  
'biscuits': 589,  
'did': 13510,  
'make': 13520,  
'often': 1837,  
'enough': 5867,  
'really': 17285,  
'mother': 765,  
'grandmother': 167,  
'used': 10184,  
'then': 9791,  
'betty': 161,  
'crocker': 136,  
'bisquick': 89,  
'mix': 7690,  
'wow': 838,  
'difference': 1949.

DIFFERENCE : 1515,  
'made': 9110,  
'rolled': 272,  
'method': 378,  
'drop': 512,  
'easy': 5832,  
'quick': 2350,  
'consistently': 259,  
'true': 1227,  
'southern': 193,  
'biscuit': 193,  
'happy': 4381,  
'man': 382,  
'his': 4620,  
'sausage': 308,  
'gravy': 464,  
'&': 7575,  
'evident': 50,  
'product.the': 3,  
'big': 4505,  
'needs.would': 1,  
'again': 9214,  
'cat': 4407,  
'crazy': 850,  
'he': 10882,  
'accepting': 11,  
'three': 3113,  
'him': 3221,  
'sensitive': 727,  
'stomach': 1297,  
'causes': 193,  
'problems': 1531,  
'only': 17071,  
'place': 1781,  
'able': 3060,  
'discontinued': 305,  
'small': 5536,  
'maintain': 182,  
'inventory': 50,  
'reason': 1798,  
'grain-free': 157,  
'grateful': 141,  
'source': 857,  
'rescued': 57,  
'sonny': 8,  
'thanks': 1794,  
'am': 12775,  
'adicated': 1,  
'say.i': 2,  
'main': 656,  
'giving': 1608,  
'5': 4515,  
'stars': 2888,  
'parmesan': 222,  
'cheese': 3338,  
'melts': 272,  
'come': 3447,  
'though': 5336,  
'hardly': 350,  
'carries': 399,  
'anymore': 914.

'see': 5104,  
'detracts': 12,  
'their': 9915,  
'jar': 1825,  
'sales': 175,  
'another': 4511,  
'cheap': 1231,  
'brick': 84,  
'mortar': 45,  
'retailer': 90,  
'': 3775,  
'places': 367,  
'hope': 1512,  
'continue': 1267,  
'kraft': 4,  
'italian': 742,  
'style': 710,  
'dinner': 1205,  
'similar': 1660,  
'same': 6423,  
'first': 10507,  
'thought': 4968,  
'after': 11837,  
'reading': 1051,  
'reviews': 3271,  
'here': 5788,  
'meal': 2590,  
'under': 1250,  
'bucks': 429,  
'cleanup': 49,  
'packet': 1319,  
'stretched': 14,  
'dinners': 98,  
'must': 2401,  
'use': 15951,  
'once': 3592,  
'wife': 1731,  
'looking': 4760,  
'cardboard': 556,  
'she': 10585,  
'lose': 534,  
'10': 2415,  
'pounds': 829,  
'should': 4508,  
'50.': 12,  
'watching': 439,  
'sugars': 494,  
'store': 7974,  
'bought': 8953,  
'foods': 3757,  
'lower': 1058,  
'sugar': 9995,  
'122': 3,  
'102': 5,  
'prediabetic': 3,  
'almost': 4167,  
'normal': 1145,  
'course': 1782,  
'disagree': 104,  
'anyone': 2339,



-  
'referring': 31,  
'anything': 4058,  
'being': 4818,  
"normal": 13,  
'crisp': 671,  
'cookies': 4386,  
'pretty': 4800,  
'granular': 38,  
'slight': 684,  
'aftertaste': 1593,  
'compared': 1607,  
'stack': 72,  
'well': 12448,  
'those': 5191,  
'trying': 3683,  
'watch': 627,  
'fats': 264,  
'measure': 256,  
'quite': 4456,  
'products': 5459,  
'given': 1406,  
'difficulty': 131,  
'finding': 981,  
'locally': 1002,  
'bottle': 3710,  
'break': 1092,  
'transit': 93,  
'considering': 680,  
'4': 4404,  
'cases': 523,  
'suppose': 385,  
'bad': 5244,  
'delivery': 1372,  
'likely': 666,  
'tastes': 8216,  
'bland': 1027,  
'unsweetened': 368,  
'lemonade': 338,  
'orange': 1656,  
'blending': 90,  
'--': 7409,  
'50/50': 59,  
'tang': 172,  
'powder': 2637,  
'zing': 153,  
'put': 5511,  
'nonfat': 59,  
'frothy': 43,  
'milk': 6778,  
'cinnamon': 2117,  
'start': 1673,  
'hearty': 296,  
'holds': 353,  
'own': 3753,  
'half': 3666,  
'complaint': 711,  
'sealed': 696,  
'plastic': 2033,  
'retain': 68,  
'freshness': 424,

'goes': 1921,  
'fast': 2340,  
'sooo': 173,  
'spokesperson': 5,  
'huge': 1559,  
'fan': 2174,  
'know': 7250,  
'drinks': 1894,  
'exactlly': 1,  
'say': 6267,  
'sleep': 489,  
'definitely': 5080,  
'go-to': 123,  
'home': 3421,  
'late': 507,  
'early': 477,  
'time': 14365,  
'unwind': 18,  
'into': 6990,  
'bed': 472,  
'sip': 469,  
'30': 1042,  
'later': 1331,  
'ready': 947,  
'slip': 86,  
'coma': 11,  
'unlike': 932,  
'aids': 39,  
'groggy': 17,  
'nights': 154,  
'bliss': 62,  
'feels': 446,  
'gas': 611,  
'dentist': 53,  
'relaxing': 143,  
'smiley': 10,  
'lean': 162,  
'sonic': 20,  
'dogs': 4845,  
'awhile': 396,  
'kept': 814,  
'...': 19399,  
'new': 4371,  
'kind': 3097,  
'snack': 5001,  
'ad': 143,  
'chips': 2962,  
'read': 2425,  
'dozens': 71,  
'reviews-': 4,  
'maybe': 2821,  
'100': 1784,  
'more.': 184,  
'common': 372,  
'comment': 380,  
'low': 3636,  
'calories': 4176,  
'were.': 10,  
'received': 3583,  
'surprised': 1608,

'3': 5588,  
'ounce': 1270,  
'360': 23,  
'calories.': 98,  
'120': 166,  
'versus': 171,  
'140': 175,  
'corn': 2672,  
'kettle': 287,  
'cooked': 1163,  
'potato': 1119,  
'chips.': 39,  
'exactly': 1756,  
'call': 1016,  
'calorie': 1145,  
'food.': 266,  
'commonly': 98,  
'sited-': 1,  
'barbecue': 129,  
'vinegar-': 1,  
'coming': 806,  
'top.': 37,  
'quality': 5752,  
'useful': 216,  
'fine': 2822,  
'texture': 4174,  
'off': 4960,  
'disappointed': 2821,  
'pamela': 306,  
'hummus': 184,  
'eat': 10917,  
'convenient': 1393,  
'lightweight': 45,  
'packed': 1164,  
'anywhere': 1021,  
'car': 495,  
'backpacking': 57,  
'picnicking': 2,  
'survival': 23,  
'kit': 557,  
'fun': 1016,  
'gluten': 3383,  
'free': 6857,  
'rice/seed': 1,  
'crackers': 2051,  
'7': 1063,  
'tubie': 1,  
'surprisingly': 362,  
'roasted': 948,  
'salted': 274,  
'soybeans': 53,  
'hot/sweet': 2,  
'nut': 1032,  
'dried': 1947,  
'cranberries': 290,  
'43': 3,  
'oz': 2580,  
'60': 434,  
'%': 4327,  
'cacao': 247,

'dark': 3464,  
'chocolate': 11135,  
'bar': 3053,  
'sturdy': 205,  
'box': 8751,  
'inside': 1490,  
'sudoku': 7,  
'puzzle': 39,  
'worth': 4096,  
'money': 3055,  
'subscribe': 1457,  
'save': 2720,  
'expiration': 742,  
'date': 1462,  
'unit': 222,  
'purchased': 4095,  
'march': 118,  
'2011': 185,  
'august': 108,  
'13': 317,  
'2011.': 71,  
'agree': 860,  
'throughout': 357,  
'lunch': 1202,  
'okay': 1041,  
'yes': 1564,  
'granola': 1384,  
'bars': 3505,  
'yummy': 1856,  
'getting': 3458,  
'ripped': 164,  
'walmart': 582,  
'packs': 1472,  
'mixes': 1134,  
'easily': 1802,  
'shaken': 35,  
'adding': 1663,  
'recommended': 2371,  
'tablespoons': 470,  
'water': 10678,  
'milkier': 2,  
'helped': 636,  
'disguise': 23,  
'dry': 3674,  
'powdered': 639,  
'drunk': 81,  
'added': 3820,  
'label': 1384,  
'twice': 1129,  
'allowed': 220,  
'chill': 96,  
'reconstituted': 52,  
'minimal': 205,  
'after-taste': 129,  
'highly': 4138,  
'wish': 2586,  
'disaster': 58,  
'preparation': 235,  
'long': 5247,  
'term': 170,  
'...': 222

'storage': 329,  
'foods/beverages': 2,  
'actually': 4579,  
'eating/drinking': 5,  
'rather': 2342,  
'known': 724,  
'friend': 1553,  
'offered': 701,  
'went': 2569,  
'hatch': 21,  
'coins': 41,  
'soaking': 123,  
'disgusting': 385,  
'yourself': 1033,  
'favor': 343,  
'steer': 63,  
'clear': 1099,  
'nasty': 616,  
'bags': 4635,  
'open': 2401,  
'line': 1572,  
'17': 241,  
'year': 3742,  
'likes': 1702,  
'olds': 47,  
'lunches': 205,  
'definatly': 293,  
'using': 5771,  
'substitutue': 14,  
'stop': 1039,  
'wo': 3210,  
'restock': 32,  
'item': 3167,  
'ok': 1496,  
'expected': 1557,  
'didnt': 258,  
'every': 6338,  
'four': 1588,  
'five': 1320,  
'six': 970,  
'seven': 207,  
'..if': 6,  
'wet': 728,  
'seemed': 1196,  
'tasted': 4372,  
'look': 2951,  
'polite': 25,  
'response': 242,  
'refund': 392,  
'package': 4637,  
'shipping': 4724,  
'try': 11985,  
'loved': 3489,  
'peanut': 3017,  
'butter': 4564,  
'cereals': 665,  
'tending': 5,  
'ridiculously': 100,  
'grew': 625,  
'eating': 4954,  
'mccccc': 115

'reese': 115,  
'cereal': 3910,  
'buying': 4957,  
'barbara': 45,  
'puffins': 11,  
'pb': 246,  
're': 7027,  
'sweetness': 1300,  
'akin': 59,  
'kix': 16,  
'd': 4730,  
'dessert': 569,  
'combination': 900,  
'reviewers': 858,  
'wrote': 238,  
'review': 2992,  
'especially': 3151,  
'dog': 8554,  
'talk': 185,  
'tell': 2029,  
'appreciate': 613,  
'leathery': 18,  
'contain': 1157,  
'papain': 4,  
'assist': 27,  
'digesting': 45,  
'hide': 279,  
'whenever': 554,  
'takes': 1397,  
'possession': 7,  
'eagerly': 64,  
'usually': 3528,  
'middle': 484,  
'll': 6036,  
'back': 5724,  
'occasion': 201,  
'bone': 545,  
'chewing': 791,  
'willingly': 24,  
'surrendered': 1,  
'chop': 92,  
'smells': 1413,  
'nothing': 3063,  
'pineapple': 371,  
'stong': 18,  
'bitter': 2715,  
'chemical': 612,  
'total': 1051,  
'waste': 1228,  
'raspberry': 680,  
'cheesecake': 111,  
'net': 181,  
'14': 367,  
'carbs': 891,  
'1': 4572,  
'barely': 490,  
'supermarket': 932,  
'tend': 691,  
'12': 2250,  
'minimum': 147,  
'turned': 1044

cupped': 1044,  
'african': 120,  
'nectar': 223,  
'son': 1981,  
'us': 3327,  
'everyday': 726,  
'mighty': 98,  
'leaf': 751,  
'brand': 6307,  
'pricey': 866,  
'pots': 138,  
'single': 1559,  
'herbal': 563,  
'teas': 2612,  
'running': 659,  
'rooibos': 295,  
'leaves': 1553,  
'combined': 295,  
'herbs': 449,  
'makes': 6474,  
'guittard': 12,  
'bring': 949,  
'recipe': 2064,  
'help': 2165,  
'straight': 845,  
'cherry': 989,  
'chip': 875,  
'grown': 465,  
'touch': 821,  
'amount': 3733,  
'sprinkle': 447,  
'top': 3041,  
'bread': 3145,  
'mini': 496,  
'muffins': 428,  
'treat': 4123,  
'enjoyed': 1827,  
'canadian': 57,  
'pancake': 322,  
'house': 2457,  
'waiting': 404,  
'sunday': 101,  
'states': 631,  
'sadly': 286,  
'purchase': 4311,  
'organic': 5832,  
'baby': 2201,  
'formula': 1577,  
'felt': 1028,  
'awful': 821,  
'breast': 313,  
'production': 188,  
'suddenly': 126,  
'came': 3713,  
'end': 1803,  
'wanted': 2510,  
'avoid': 904,  
'months': 3169,  
'canister': 255,  
'similac': 156,  
'b/c': 186.

```

~,~': 100,
'discovered': 1091,
'earths': 32,
'yet': 2400,
'spit': 261,
'lot': 6224,
'each': 4840,
'feeding': 988,
'tomorrow': 86,
'18': 452,
'pound': 1164,
'boy': 492,
'pleased': 1636,
'forward': 849,
'offers': 412,
'solids': 132,
'hazelnut': 623,
'balanced': 399,
'toffee': 225,
'sounds': 321,
'else': 2268,
'?: 9340,
'gim': 17,
'pickled': 79,
'besides': 381,
'sushi': 239,
'snacking': 307,
'comes': 3194,
'looks': 1476,
'restaurants': 364,
'thin': 1044,
'slices': 370,
'flavor.': 483,
'aspartame': 277,
'equal': 357,
'sweeter': 563,
'need': 5489,
'glucose': 137,
'levels': 401,
'unfortunately': 1429,
...})

```

In [118]:

```

top_words = pd.DataFrame(word_dict.most_common(nu_top_words), columns= ['words', 'frequency'])

```

In [119]:

```

top_words

```

Out[119]:

	words	frequency
0	.	420578
1	the	327317
2	,	304782
3	:	205160



3	1	295100
	<b>words</b>	<b>frequency</b>
4	and	224509
5	a	212402
6	it	191898
7	to	177086
8	of	138804
9	is	132579
10	this	120930
11	>	119781
12	<	119401
13	/	114085
14	br	113681
15	in	97932
16	for	95763
17	that	77976
18	my	77680
19	!	75535
20	you	65155
21	but	64539
22	with	62737
23	have	61096
24	not	60013
25	was	58367
26	are	54258
27	they	52252
28	's	52130
29	n't	51722
...	...	...
4970	thirsty	72
4971	salt.	72
4972	ms	72
4973	bittersweet	72
4974	wilderness	72
4975	dozens	71
4976	2011.	71
4977	respond	71

	words	frequency
4978	explained	71
4979	lemony	71
4980	fortified	71
4981	cottage	71
4982	problem.	71
4983	deeper	71
4984	believed	71
4985	ideas	71
4986	gripe	71
4987	whats	71
4988	noses	71
4989	curiosity	71
4990	stainless	71
4991	exceeded	71
4992	oregano	71
4993	gray	71
4994	slightest	71
4995	phenomenal	71
4996	5-star	71
4997	b6	71
4998	whisk	71
4999	steamer	71

5000 rows × 2 columns

In [120]:

```
top_dict = {}
for idx,tup in enumerate(word_dict.most_common(nu_top_words)):
    top_dict[tup[0]] = idx
```

In [121]:

```
top_dict
```

Out[121]:

```
{'.': 0,
 'the': 1,
 ',': 2,
 'i': 3,
 'and': 4,
 'a': 5,
 'it': 6,
 'to': 7,
```

'of': 8,  
'is': 9,  
'this': 10,  
'>': 11,  
'<': 12,  
'/': 13,  
'br': 14,  
'in': 15,  
'for': 16,  
'that': 17,  
'my': 18,  
'!': 19,  
'you': 20,  
'but': 21,  
'with': 22,  
'have': 23,  
'not': 24,  
'was': 25,  
'are': 26,  
'they': 27,  
"'s": 28,  
"n't": 29,  
'as': 30,  
'on': 31,  
'so': 32,  
'like': 33,  
')': 34,  
'these': 35,  
'(': 36,  
'good': 37,  
'them': 38,  
'be': 39,  
'or': 40,  
'at': 41,  
'just': 42,  
'taste': 43,  
'very': 44,  
'if': 45,  
'one': 46,  
'do': 47,  
'great': 48,  
'product': 49,  
'can': 50,  
'all': 51,  
'from': 52,  
"''": 53,  
'flavor': 54,  
'coffee': 55,  
'we': 56,  
'tea': 57,  
'when': 58,  
'would': 59,  
'more': 60,  
'has': 61,  
'had': 62,  
'will': 63,  
'me': 64,  
'love': 65,  
'than': 66,  
'out': 67,

'...': 68,  
'``': 69,  
'other': 70,  
'no': 71,  
'get': 72,  
'some': 73,  
'about': 74,  
'really': 75,  
'only': 76,  
'up': 77,  
'food': 78,  
'much': 79,  
'use': 80,  
'an': 81,  
'too': 82,  
'there': 83,  
'your': 84,  
' ': 85,  
'also': 86,  
" 've ": 87,  
'were': 88,  
'-': 89,  
'amazon': 90,  
'little': 91,  
'because': 92,  
'time': 93,  
'what': 94,  
'which': 95,  
'make': 96,  
'buy': 97,  
'did': 98,  
'best': 99,  
'price': 100,  
'find': 101,  
'been': 102,  
" 'm ": 103,  
'even': 104,  
'does': 105,  
'am': 106,  
'tried': 107,  
'by': 108,  
'well': 109,  
'better': 110,  
'try': 111,  
'after': 112,  
'now': 113,  
'any': 114,  
'chocolate': 115,  
'eat': 116,  
'he': 117,  
'water': 118,  
'she': 119,  
'first': 120,  
'used': 121,  
'could': 122,  
'sugar': 123,  
'their': 124,  
'then': 125,  
'found': 126,  
'our': 127,

'?': 128,  
'again': 129,  
'made': 130,  
'sweet': 131,  
'drink': 132,  
'bag': 133,  
'bought': 134,  
'box': 135,  
'way': 136,  
'dog': 137,  
'think': 138,  
'tastes': 139,  
'two': 140,  
'over': 141,  
'who': 142,  
'store': 143,  
'since': 144,  
'cup': 145,  
'still': 146,  
'order': 147,  
'mix': 148,  
'\$': 149,  
'&': 150,  
'most': 151,  
'bit': 152,  
'delicious': 153,  
'her': 154,  
'day': 155,  
'nice': 156,  
'go': 157,  
'--': 158,  
'recommend': 159,  
'2': 160,  
'many': 161,  
'got': 162,  
'know': 163,  
'add': 164,  
'give': 165,  
're': 166,  
'into': 167,  
'how': 168,  
'want': 169,  
'ca': 170,  
'before': 171,  
'free': 172,  
'hot': 173,  
'milk': 174,  
'never': 175,  
'stuff': 176,  
'something': 177,  
'favorite': 178,  
'makes': 179,  
'few': 180,  
';': 181,  
'same': 182,  
'without': 183,  
'every': 184,  
'years': 185,  
'brand': 186,  
'flavors': 187,

'say': 188,  
'while': 189,  
'its': 190,  
'lot': 191,  
'always': 192,  
'right': 193,  
"ll": 194,  
'different': 195,  
'less': 196,  
'enough': 197,  
'ever': 198,  
'easy': 199,  
'organic': 200,  
'however': 201,  
'here': 202,  
'using': 203,  
'quality': 204,  
'ordered': 205,  
'back': 206,  
'sauce': 207,  
'perfect': 208,  
'fresh': 209,  
'3': 210,  
'keep': 211,  
'oil': 212,  
'healthy': 213,  
'small': 214,  
'put': 215,  
'ingredients': 216,  
'need': 217,  
'sure': 218,  
'products': 219,  
'whole': 220,  
'though': 221,  
'loves': 222,  
'long': 223,  
'bad': 224,  
'those': 225,  
'hard': 226,  
'pack': 227,  
'see': 228,  
'local': 229,  
'definitely': 230,  
'enjoy': 231,  
'green': 232,  
'snack': 233,  
'thought': 234,  
'off': 235,  
'buying': 236,  
'eating': 237,  
'regular': 238,  
'old': 239,  
'strong': 240,  
'dogs': 241,  
'each': 242,  
'being': 243,  
'pretty': 244,  
'far': 245,  
'thing': 246,  
'looking': 247,  
..

'both': 248,  
"d": 249,  
'shipping': 250,  
'high': 251,  
'rice': 252,  
'work': 253,  
'package': 254,  
'bags': 255,  
'down': 256,  
'his': 257,  
'tasty': 258,  
'actually': 259,  
'1': 260,  
'butter': 261,  
'take': 262,  
'size': 263,  
'going': 264,  
'5': 265,  
'another': 266,  
'should': 267,  
'big': 268,  
'quite': 269,  
'natural': 270,  
'salt': 271,  
'through': 272,  
'last': 273,  
'people': 274,  
'cat': 275,  
'4': 276,  
'wonderful': 277,  
'cookies': 278,  
'chicken': 279,  
'happy': 280,  
'tasted': 281,  
'new': 282,  
'%': 283,  
'purchase': 284,  
'may': 285,  
'excellent': 286,  
'treats': 287,  
'calories': 288,  
'texture': 289,  
'almost': 290,  
'highly': 291,  
'real': 292,  
'treat': 293,  
'worth': 294,  
'purchased': 295,  
'per': 296,  
'anything': 297,  
'stores': 298,  
'candy': 299,  
'grocery': 300,  
'around': 301,  
'coconut': 302,  
'cereal': 303,  
'feel': 304,  
'added': 305,  
"": 306,  
'foods': 307,  
'candy': 308

'fruit': 308,  
'own': 309,  
'tasting': 310,  
'year': 311,  
'amount': 312,  
'came': 313,  
'family': 314,  
'bottle': 315,  
'having': 316,  
'several': 317,  
'trying': 318,  
'probably': 319,  
'dry': 320,  
'expensive': 321,  
'half': 322,  
'diet': 323,  
'low': 324,  
'once': 325,  
'received': 326,  
'said': 327,  
'usually': 328,  
'vanilla': 329,  
'bars': 330,  
'loved': 331,  
'fat': 332,  
'brands': 333,  
'dark': 334,  
'getting': 335,  
'come': 336,  
'home': 337,  
'why': 338,  
'smell': 339,  
'making': 340,  
'full': 341,  
'gluten': 342,  
'company': 343,  
'where': 344,  
'away': 345,  
'might': 346,  
'cheese': 347,  
'things': 348,  
'us': 349,  
'6': 350,  
'days': 351,  
'reviews': 352,  
'arrived': 353,  
'black': 354,  
'protein': 355,  
'him': 356,  
'beans': 357,  
'wo': 358,  
'morning': 359,  
'comes': 360,  
'kids': 361,  
'available': 362,  
'case': 363,  
'cats': 364,  
'months': 365,  
'item': 366,  
'light': 367,  
'especially': 368



especially': 368,  
'bread': 369,  
'seems': 370,  
'three': 371,  
'kind': 372,  
'until': 373,  
'nothing': 374,  
'able': 375,  
'problem': 376,  
'such': 377,  
'money': 378,  
'bar': 379,  
'top': 380,  
'honey': 381,  
'peanut': 382,  
'drinking': 383,  
'review': 384,  
'instead': 385,  
'large': 386,  
'chips': 387,  
'health': 388,  
'look': 389,  
'blend': 390,  
'least': 391,  
'stars': 392,  
'husband': 393,  
'http': 394,  
'soup': 395,  
'fine': 396,  
'maybe': 397,  
'disappointed': 398,  
'juice': 399,  
'times': 400,  
'href=': 401,  
'/a': 402,  
'pasta': 403,  
'boxes': 404,  
'extra': 405,  
'minutes': 406,  
'flavored': 407,  
'cream': 408,  
'packaging': 409,  
'white': 410,  
'save': 411,  
'others': 412,  
'next': 413,  
'fact': 414,  
'bitter': 415,  
'ones': 416,  
'let': 417,  
'corn': 418,  
'either': 419,  
'syrup': 420,  
'gave': 421,  
'works': 422,  
'variety': 423,  
'powder': 424,  
'prefer': 425,  
'breakfast': 426,  
'teas': 427,  
'decided': 428.

accused': 428,  
'liked': 429,  
'deal': 430,  
'absolutely': 431,  
'couple': 432,  
'meal': 433,  
'wish': 434,  
'oz': 435,  
'went': 436,  
'although': 437,  
'hair': 438,  
'cheaper': 439,  
'cost': 440,  
'smooth': 441,  
'wanted': 442,  
'red': 443,  
'soft': 444,  
'started': 445,  
'ago': 446,  
'quickly': 447,  
'house': 448,  
'serving': 449,  
'plus': 450,  
'energy': 451,  
'read': 452,  
'10': 453,  
'must': 454,  
'open': 455,  
'yet': 456,  
'cans': 457,  
'popcorn': 458,  
'recommended': 459,  
'rich': 460,  
'myself': 461,  
'longer': 462,  
'quick': 463,  
'gift': 464,  
'rather': 465,  
'fast': 466,  
'anyone': 467,  
'everything': 468,  
'seem': 469,  
'ordering': 470,  
'took': 471,  
'amazing': 472,  
'else': 473,  
'cups': 474,  
'12': 475,  
'pieces': 476,  
'glad': 477,  
'cookie': 478,  
'spicy': 479,  
'baby': 480,  
'ginger': 481,  
'fan': 482,  
'cooking': 483,  
'help': 484,  
'cinnamon': 485,  
'ice': 486,  
'wheat': 487,  
'week': 488.

'cold': 489,  
'side': 490,  
'meat': 491,  
'everyone': 492,  
'recipe': 493,  
'mouth': 494,  
'crackers': 495,  
'second': 496,  
'mixed': 497,  
'plastic': 498,  
'tell': 499,  
'friends': 500,  
'roast': 501,  
'month': 502,  
'part': 503,  
'sometimes': 504,  
'flour': 505,  
'son': 506,  
'cook': 507,  
'weight': 508,  
'version': 509,  
'etc': 510,  
'difference': 511,  
'fiber': 512,  
'dried': 513,  
'container': 514,  
'nuts': 515,  
'thank': 516,  
'noodles': 517,  
'goes': 518,  
'market': 519,  
'says': 520,  
'brown': 521,  
'believe': 522,  
'drinks': 523,  
'slightly': 524,  
'online': 525,  
'value': 526,  
'super': 527,  
'overall': 528,  
'yummy': 529,  
'crunchy': 530,  
'heat': 531,  
'often': 532,  
'enjoyed': 533,  
'jar': 534,  
'oatmeal': 535,  
'type': 536,  
'gum': 537,  
'opened': 538,  
'spice': 539,  
'soy': 540,  
'end': 541,  
'easily': 542,  
'awesome': 543,  
'reason': 544,  
'thanks': 545,  
'cake': 546,  
'8': 547,  
'100': 548,

'course': 549,  
'place': 550,  
'life': 551,  
'artificial': 552,  
'lemon': 553,  
'guess': 554,  
'cut': 555,  
'exactly': 556,  
'daughter': 557,  
'gets': 558,  
'bowl': 559,  
'wife': 560,  
'plain': 561,  
'color': 562,  
'hand': 563,  
'care': 564,  
'left': 565,  
'seeds': 566,  
'likes': 567,  
'sodium': 568,  
'smaller': 569,  
'choice': 570,  
'chew': 571,  
'start': 572,  
'original': 573,  
'starbucks': 574,  
'beef': 575,  
'canned': 576,  
'ingredient': 577,  
'adding': 578,  
'similar': 579,  
'orange': 580,  
'bottles': 581,  
'recently': 582,  
'pepper': 583,  
'pleased': 584,  
'pay': 585,  
'between': 586,  
'machine': 587,  
'ground': 588,  
'cocoa': 589,  
'soda': 590,  
'special': 591,  
'contains': 592,  
'weeks': 593,  
'alternative': 594,  
'giving': 595,  
'surprised': 596,  
'compared': 597,  
'itself': 598,  
'wrong': 599,  
'aftertaste': 600,  
'snacks': 601,  
'apple': 602,  
'brew': 603,  
'daily': 604,  
'four': 605,  
'salty': 606,  
'simply': 607,  
'caffeine': 608,

'someone': 609,  
'french': 610,  
'stick': 611,  
'formula': 612,  
'line': 613,  
'experience': 614,  
'during': 615,  
'yes': 616,  
'already': 617,  
'huge': 618,  
'single': 619,  
'expected': 620,  
'instant': 621,  
'friend': 622,  
'leaves': 623,  
'mild': 624,  
'ate': 625,  
'saw': 626,  
'live': 627,  
'past': 628,  
'clean': 629,  
'looked': 630,  
'aroma': 631,  
'along': 632,  
'problems': 633,  
'close': 634,  
'fish': 635,  
'flavorful': 636,  
'espresso': 637,  
'run': 638,  
'lots': 639,  
'baking': 640,  
'hope': 641,  
'teeth': 642,  
'ok': 643,  
'list': 644,  
'inside': 645,  
'night': 646,  
'idea': 647,  
'packaged': 648,  
'looks': 649,  
'packs': 650,  
'change': 651,  
'keurig': 652,  
'finally': 653,  
'date': 654,  
'needed': 655,  
'fantastic': 656,  
'subscribe': 657,  
'glass': 658,  
'decaf': 659,  
'raw': 660,  
'service': 661,  
'unfortunately': 662,  
'noticed': 663,  
'shipped': 664,  
'olive': 665,  
'smells': 666,  
'gone': 667,  
'due': 668,

'given': 669,  
'star': 670,  
'hours': 671,  
'k-cups': 672,  
'takes': 673,  
'gives': 674,  
'convenient': 675,  
'bite': 676,  
'iced': 677,  
'bold': 678,  
'done': 679,  
'granola': 680,  
'label': 681,  
'coffees': 682,  
'20': 683,  
'items': 684,  
'bulk': 685,  
'chewy': 686,  
'filling': 687,  
'stock': 688,  
'it.': 689,  
'delivery': 690,  
'pot': 691,  
'creamy': 692,  
'christmas': 693,  
'mine': 694,  
'packets': 695,  
'extremely': 696,  
'helps': 697,  
'completely': 698,  
'later': 699,  
'eaten': 700,  
'needs': 701,  
'liquid': 702,  
'sent': 703,  
'bottom': 704,  
'sold': 705,  
'five': 706,  
'packet': 707,  
'packages': 708,  
'leave': 709,  
'1/2': 710,  
'larger': 711,  
'soon': 712,  
'recipes': 713,  
'expect': 714,  
'sweetness': 715,  
'carry': 716,  
'within': 717,  
'stomach': 718,  
'please': 719,  
'pet': 720,  
'rest': 721,  
'mind': 722,  
'feed': 723,  
'salad': 724,  
'body': 725,  
'grain': 726,  
'wait': 727,  
'ounce': 728,

'opinion': 729,  
'thick': 730,  
'continue': 731,  
'under': 732,  
'picky': 733,  
'sour': 734,  
'today': 735,  
'name': 736,  
'certainly': 737,  
'blue': 738,  
'cheap': 739,  
'easier': 740,  
'waste': 741,  
'true': 742,  
'note': 743,  
'called': 744,  
'anyway': 745,  
'excited': 746,  
'grams': 747,  
'skin': 748,  
'garlic': 749,  
'seller': 750,  
'based': 751,  
'world': 752,  
'together': 753,  
'bean': 754,  
'vet': 755,  
'dinner': 756,  
'pods': 757,  
'lunch': 758,  
'seemed': 759,  
'point': 760,  
'microwave': 761,  
'doing': 762,  
'almond': 763,  
'truly': 764,  
'tuna': 765,  
'chai': 766,  
'chili': 767,  
'packed': 768,  
'pound': 769,  
'cooked': 770,  
'addition': 771,  
'content': 772,  
'difficult': 773,  
'contain': 774,  
'normally': 775,  
'purchasing': 776,  
'healthier': 777,  
'vitamin': 778,  
'tiny': 779,  
'seasoning': 780,  
'remember': 781,  
'dish': 782,  
'normal': 783,  
'calorie': 784,  
'sweetener': 785,  
'consistency': 786,  
'seen': 787,  
'told': 788,  
'...': 789

'set': 789,  
'pure': 790,  
'entire': 791,  
'simple': 792,  
'jerky': 793,  
'mixes': 794,  
'stevia': 795,  
'twice': 796,  
'area': 797,  
'spices': 798,  
'potato': 799,  
'hint': 800,  
'batch': 801,  
'benefits': 802,  
'[': 803,  
'decent': 804,  
'delivered': 805,  
']': 806,  
'unless': 807,  
'refreshing': 808,  
'clear': 809,  
'k-cup': 810,  
'become': 811,  
'break': 812,  
'discovered': 813,  
'substitute': 814,  
'worked': 815,  
'15': 816,  
'licorice': 817,  
'almonds': 818,  
'maker': 819,  
'customer': 820,  
'particular': 821,  
'7': 822,  
'results': 823,  
'medium': 824,  
'perhaps': 825,  
'lower': 826,  
'sale': 827,  
'issues': 828,  
'reading': 829,  
'total': 830,  
'piece': 831,  
'baked': 832,  
'strawberry': 833,  
'turned': 834,  
'thin': 835,  
'oh': 836,  
'30': 837,  
'okay': 838,  
'stop': 839,  
'mint': 840,  
'yourself': 841,  
'nut': 842,  
'taking': 843,  
'hit': 844,  
'felt': 845,  
'immediately': 846,  
'bland': 847,  
'option': 848,  
'stopped': 849



'stopped': 849,  
'anywhere': 850,  
'means': 851,  
'including': 852,  
'call': 853,  
'fun': 854,  
'crunch': 855,  
'broken': 856,  
'totally': 857,  
'yogurt': 858,  
'stay': 859,  
'description': 860,  
'person': 861,  
'weak': 862,  
'pumpkin': 863,  
'locally': 864,  
'pop': 865,  
'loose': 866,  
'none': 867,  
'keeps': 868,  
'pleasant': 869,  
'cherry': 870,  
'nutrition': 871,  
'matter': 872,  
'feeding': 873,  
'serve': 874,  
'throw': 875,  
'higher': 876,  
'acid': 877,  
'finding': 878,  
'kitchen': 879,  
'sell': 880,  
'uses': 881,  
'thinking': 882,  
'six': 883,  
'24': 884,  
'shop': 885,  
'shipment': 886,  
'plan': 887,  
'check': 888,  
'stale': 889,  
'bring': 890,  
'roasted': 891,  
'ready': 892,  
'mostly': 893,  
'perfectly': 894,  
'website': 895,  
'wine': 896,  
'warm': 897,  
'satisfying': 898,  
'issue': 899,  
'sort': 900,  
'caramel': 901,  
'unlike': 902,  
'supermarket': 903,  
'varieties': 904,  
'amazon.com': 905,  
'eggs': 906,  
'peanuts': 907,  
'whatever': 908,  
'homemade': 909

homemade': 909,  
'nearly': 910,  
'feeling': 911,  
'short': 912,  
'reasonable': 913,  
'listed': 914,  
'anymore': 915,  
'condition': 916,  
'heavy': 917,  
'finish': 918,  
'jars': 919,  
'oats': 920,  
'somewhat': 921,  
'sauces': 922,  
'reviewer': 923,  
'avoid': 924,  
'sticks': 925,  
'otherwise': 926,  
'combination': 927,  
'eats': 928,  
'pour': 929,  
'pick': 930,  
'except': 931,  
'mess': 932,  
'carbs': 933,  
'gluten-free': 934,  
'notice': 935,  
'salmon': 936,  
'50': 937,  
'supposed': 938,  
'paid': 939,  
'hoping': 940,  
'future': 941,  
'chip': 942,  
'instructions': 943,  
'moist': 944,  
'horrible': 945,  
'mom': 946,  
'adds': 947,  
'impressed': 948,  
'pricey': 949,  
'expecting': 950,  
'number': 951,  
'summer': 952,  
'terrible': 953,  
'changed': 954,  
'agree': 955,  
'types': 956,  
'near': 957,  
'veggies': 958,  
'reviewers': 959,  
'source': 960,  
'directions': 961,  
'stronger': 962,  
'crazy': 963,  
'forward': 964,  
'fairly': 965,  
'straight': 966,  
'beat': 967,  
'prices': 968,  
'shelf': 969.

```

'smell': 969,
'brewed': 970,
'china': 971,
'splenda': 972,
'wow': 973,
'picture': 974,
'yum': 975,
'nutritional': 976,
'vegetables': 977,
'office': 978,
'pounds': 979,
'chews': 980,
'hands': 981,
'candies': 982,
'puppy': 983,
'individual': 984,
'touch': 985,
'awful': 986,
'satisfied': 987,
'kept': 988,
'switch': 989,
'fit': 990,
'gourmet': 991,
'return': 992,
'shampoo': 993,
'ended': 994,
'children': 995,
'vitamins': 996,
'dont': 997,
'hooked': 998,
'coming': 999,
...}

```

In [122]:

```

for idx,row in final.iterrows():
    words = nltk.tokenize.word_tokenize(row['Text'])
    words_list = []
    for word in words:
        if word in top_dict.keys():
            words_list.append(top_dict.get(word))
    final.set_value(idx,'Text',words_list)

```

C:\Users\Anil Chowdary\Anaconda3\lib\site-packages\ipykernel\_launcher.py:7:  
FutureWarning: set\_value is deprecated and will be removed in a future release. Please use .at[] or .iat[] accessors instead  
import sys

In [123]:

```
final.head(5)
```

Out[123]:

	Text	Score
<b>460022</b>	[36, 3731, 4475, 923, 34, 12, 14, 13, 11, 12, ...	positive
<b>544096</b>	[3, 554, 3, 103, 24, 1993, 1, 70, 18, 363, 30,...	negative

<b>564627</b>	[56, 445, 776, 1769, 3331, 55, 74, 1604, 185, ...]	<b>Text</b>	<b>Score</b>
<b>403438</b>	[35, 26, 543, 4, 27, 47, 29, 23, 1, 1003, 251, ...]		positive
<b>8842</b>	[61, 7, 39, 1, 1179, 55, 3, 87, 198, 62, 0, 6, ...]		negative

In [124]:

```
final.shape
```

Out[124]:

```
(100000, 2)
```

In [159]:

```
final['Score'] = final['Score'].replace('positive',1)
```

In [160]:

```
final['Score'] = final['Score'].replace('negative',0)
```

In [161]:

```
final['Score']
```

Out[161]:

```
460022      1
544096      0
564627      1
403438      1
8842        0
406370      1
552938      1
304305      1
68732       1
159646      1
115320      1
566755      1
526225      0
193379      1
385442      1
448751      1
320497      1
316795      1
134606      1
381856      1
184417      1
124358      1
550509      1
150596      1
172317      1
481750      0
150064      1
428337      1
86774       1
460397      0
..
57919       1
90376       1
100000      1
```

```
43027      1
202070     1
305543     1
160428     1
74950      1
80385      0
369828     1
391425     1
473985     1
13008      1
71410      1
180680     1
226009     1
14135      1
97039      1
218920     1
113439     1
47763      1
404670     1
51315      1
374749     1
479805     0
196415     1
183667     0
44637      1
394493     1
445955     1
281361     1
Name: Score, Length: 100000, dtype: int64
```

In [162]:

```
x_train = final['Text'][0:70000]
x_test = final['Text'][70000:]
y_train = final['Score'][0:70000]
y_test = final['Score'][70000:]
```

In [163]:

```
print(x_train.shape)
print(x_test.shape)
print(y_train.shape)
print(y_test.shape)
```

```
(70000,)
(30000,)
(70000,)
(30000,)
```

In [164]:

```
from keras.preprocessing import sequence
from keras.layers import Embedding
from keras.models import Sequential
from keras.layers import LSTM
from keras.layers import Dense
```

PADDING

In [165]:

In [165]:

```
max_length = 600
x_train = sequence.pad_sequences(x_train, maxlen= max_length)
x_test = sequence.pad_sequences(x_test, maxlen= max_length)
```

In [166]:

```
print(x_train.shape)
print(x_test.shape)
```

(70000, 600)

(30000, 600)

## 1 LAYER LSTM MODEL WITH 100 LSTM CELLS

In [175]:

```
embed_vector_length = 32
model = Sequential()
model.add(Embedding(nu_top_words, embed_vector_length, input_length= max_length))
model.add(LSTM(100))
model.add(Dense(1,activation = 'sigmoid'))
model.compile(loss = 'binary_crossentropy', optimizer = 'adam', metrics = ['accuracy'])
print(model.summary())
```

Layer (type)	Output Shape	Param #
=====		
embedding_18 (Embedding)	(None, 600, 32)	160000
-----		
lstm_34 (LSTM)	(None, 100)	53200
-----		
dense_14 (Dense)	(None, 1)	101
=====		
Total params: 213,301		
Trainable params: 213,301		
Non-trainable params: 0		
-----		
None		

In [135]:

```
model.fit(x_train, y_train, batch_size= 128, nb_epoch = 15)
scores = model.evaluate(x_test, y_test)
print("Accuracy", scores[1]*100)
```

C:\Users\Anil Chowdary\Anaconda3\lib\site-packages\keras\models.py:981: UserWarning: The `nb\_epoch` argument in `fit` has been renamed `epochs`.  
warnings.warn('The `nb\_epoch` argument in `fit` '

Epoch 1/15

70000/70000 [=====] - 820s 12ms/step - loss: 0.2676 - acc: 0.8910

Epoch 2/15

70000/70000 [=====] - 787s 11ms/step - loss: 0.2455 - acc: 0.9005

Epoch 3/15

70000/70000 [=====] - 699s 10ms/step - loss: 0.2307 - acc: 0.9083

```

Epoch 4/15
70000/70000 [=====] - 703s 10ms/step - loss: 0.217
2 - acc: 0.9130
Epoch 5/15
70000/70000 [=====] - 700s 10ms/step - loss: 0.201
1 - acc: 0.9210
Epoch 6/15
70000/70000 [=====] - 698s 10ms/step - loss: 0.181
6 - acc: 0.9291
Epoch 7/15
70000/70000 [=====] - 696s 10ms/step - loss: 0.167
0 - acc: 0.9356
Epoch 8/15
70000/70000 [=====] - 701s 10ms/step - loss: 0.155
7 - acc: 0.9418
Epoch 9/15
70000/70000 [=====] - 728s 10ms/step - loss: 0.143
4 - acc: 0.9471
Epoch 10/15
70000/70000 [=====] - 771s 11ms/step - loss: 0.143
1 - acc: 0.9469
Epoch 11/15
70000/70000 [=====] - 762s 11ms/step - loss: 0.142
9 - acc: 0.9460
Epoch 12/15
70000/70000 [=====] - 799s 11ms/step - loss: 0.130
1 - acc: 0.9527
Epoch 13/15
70000/70000 [=====] - 799s 11ms/step - loss: 0.112
3 - acc: 0.9604
Epoch 14/15
70000/70000 [=====] - 723s 10ms/step - loss: 0.100
7 - acc: 0.9651
Epoch 15/15
70000/70000 [=====] - 798s 11ms/step - loss: 0.094
6 - acc: 0.9680
30000/30000 [=====] - 147s 5ms/step
Accuracy 89.15333333333334

```

### 3 LAYER LSTM MODEL WITH 100, 50 AND 25 CELLS IN THE LAYERS

In [174]:

```

embed_vector_length = 32
model = Sequential()
model.add(Embedding(nu_top_words, embed_vector_length, input_length= max_length))
model.add(LSTM(100, return_sequences= True))
model.add(LSTM(50, return_sequences= True))
model.add(LSTM(25, return_sequences= True))
model.add(Dense(1,activation = 'sigmoid'))
model.compile(loss = 'binary_crossentropy', optimizer = 'adam', metrics = ['accuracy'])
print(model.summary())

```

Layer (type)	Output Shape	Param #
embedding_17 (Embedding)	(None, 600, 32)	160000

lstm_31 (LSTM)	(None, 600, 100)	53200
lstm_32 (LSTM)	(None, 600, 50)	30200
lstm_33 (LSTM)	(None, 600, 25)	7600
dense_13 (Dense)	(None, 600, 1)	26
=====		
Total params: 251,026		
Trainable params: 251,026		
Non-trainable params: 0		
None		

In [176]:

```
model.fit(x_train, y_train, batch_size= 128, nb_epoch = 15)
scores = model.evaluate(x_test, y_test)
print("Accuracy", scores[1]*100)
```

C:\Users\Anil Chowdary\Anaconda3\lib\site-packages\keras\models.py:981: UserWarning: The `nb\_epoch` argument in `fit` has been renamed `epochs`.  
warnings.warn('The `nb\_epoch` argument in `fit` ')

```
Epoch 1/15
70000/70000 [=====] - 779s 11ms/step - loss: 0.368
7 - acc: 0.8380
Epoch 2/15
70000/70000 [=====] - 917s 13ms/step - loss: 0.272
6 - acc: 0.8874
Epoch 3/15
70000/70000 [=====] - 816s 12ms/step - loss: 0.249
2 - acc: 0.8982
Epoch 4/15
70000/70000 [=====] - 944s 13ms/step - loss: 0.232
7 - acc: 0.9057
Epoch 5/15
70000/70000 [=====] - 744s 11ms/step - loss: 0.215
5 - acc: 0.9128
Epoch 6/15
70000/70000 [=====] - 710s 10ms/step - loss: 0.203
4 - acc: 0.9192
Epoch 7/15
70000/70000 [=====] - 798s 11ms/step - loss: 0.183
5 - acc: 0.9282
Epoch 8/15
70000/70000 [=====] - 744s 11ms/step - loss: 0.170
6 - acc: 0.9350
Epoch 9/15
70000/70000 [=====] - 813s 12ms/step - loss: 0.157
4 - acc: 0.9404
Epoch 10/15
70000/70000 [=====] - 685s 10ms/step - loss: 0.145
1 - acc: 0.9462
Epoch 11/15
70000/70000 [=====] - 685s 10ms/step - loss: 0.134
0 - acc: 0.9511
Epoch 12/15
70000/70000 [=====] - 687s 10ms/step - loss: 0.123
5 - acc: 0.9558
Epoch 13/15
```



```
Epoch 13/15
70000/70000 [=====] - 688s 10ms/step - loss: 0.113
1 - acc: 0.9605
Epoch 14/15
70000/70000 [=====] - 686s 10ms/step - loss: 0.105
6 - acc: 0.9638
Epoch 15/15
70000/70000 [=====] - 685s 10ms/step - loss: 0.096
1 - acc: 0.9662
30000/30000 [=====] - 108s 4ms/step
Accuracy 88.87
```

## CONCLUSION:

- > The total length of the corpus is 9542183
- > After removing all the duplicates and cleaning all the sentences and here we considered the top 5000 words
- > Frequency of each word is calculated and indexed them based on the frequency
- > All the words in the sentences are replaced by the index positions of the top 5000 words
- > If a word in sentence is not present in top 5000 then it is discarded
- > Now we have all the reviews with the words replaced by their index positions in the frequency distribution
- > Padding all the sentences to have same length to perform batch operations

At the end of 15 epochs:

Single LSTM Layer with 100 cells:

Train loss: 0.0946                      Accuracy: 96.80

Accuracy on test data is 89.15

3 Layer LSTM model with 100, 50 and 25 cells in the consecutive layers:

Train loss: 0.0961                      Accuracy: 96.62

Accuracy on the test data is 88.87