

## word to vector using gensim

April 10, 2023

```
[2]: import gensim
import pandas as pd
```

```
[9]: df=pd.read_json(r'C:
↳\Users\Rakesh\Downloads\reviews_Cell_Phones_and_Accessories_5.
↳json\Cell_Phones_and_Accessories_5.json',lines=True)
```

```
[10]: df
```

```
[10]:
```

	reviewerID	asin	reviewerName	helpful	\
0	A30TL5EWN6DFXT	120401325X	christina	[0, 0]	
1	ASY55RVNIILOUD	120401325X	emily l.	[0, 0]	
2	A2TMXE2AF07ONB	120401325X	Erica	[0, 0]	
3	AWJOWZQYMYFQ4	120401325X	JM	[4, 4]	
4	ATX7CZYFXI1KW	120401325X	patrice m rogoza	[2, 3]	
...	...	...	...	...	
194434	A1YMNTFLNDYQ1F	BO0LORXVUE	eyeused2loveher	[0, 0]	
194435	A15TX8B2L8B2OS	BO0LORXVUE	Jon Davidson	[0, 0]	
194436	A3JI7QRZ01QG8X	BO0LORXVUE	Joyce M. Davidson	[0, 0]	
194437	A1NHB2VC68YQNM	BO0LORXVUE	Nurse Farrugia	[0, 0]	
194438	A1AG6U022WHXBF	BO0LORXVUE	Trisha Crocker	[0, 0]	

	reviewText	overall	\
0	They look good and stick good! I just don't li...	4	
1	These stickers work like the review says they ...	5	
2	These are awesome and make my phone look so st...	5	
3	Item arrived in great time and was in perfect ...	4	
4	awesome! stays on, and looks great. can be use...	5	
...	...	...	
194434	Works great just like my original one. I reall...	5	
194435	Great product. Great packaging. High quality a...	5	
194436	This is a great cable, just as good as the mor...	5	
194437	I really like it becasue it works well with my...	5	
194438	product as described, I have wasted a lot of m...	5	

	summary	unixReviewTime	\
0	Looks Good	1400630400	
1	Really great product.	1389657600	

2	LOVE LOVE LOVE	1403740800
3	Cute!	1382313600
4	leopard home button sticker for iphone 4s	1359849600
...	...	...
194434	This works just perfect!	1405900800
194435	Great replacement cable. Apple certified	1405900800
194436	Real quality	1405900800
194437	I really like it becasue it works well with my...	1405814400
194438	I have wasted a lot of money on cords	1405900800

	reviewTime
0	05 21, 2014
1	01 14, 2014
2	06 26, 2014
3	10 21, 2013
4	02 3, 2013
...	...
194434	07 21, 2014
194435	07 21, 2014
194436	07 21, 2014
194437	07 20, 2014
194438	07 21, 2014

[194439 rows x 9 columns]

```
[12]: df['reviewText']
```

```
[12]: 0      They look good and stick good! I just don't li...
      1      These stickers work like the review says they ...
      2      These are awesome and make my phone look so st...
      3      Item arrived in great time and was in perfect ...
      4      awesome! stays on, and looks great. can be use...

      ...
      194434    Works great just like my original one. I reall...
      194435    Great product. Great packaging. High quality a...
      194436    This is a great cable, just as good as the mor...
      194437    I really like it becasue it works well with my...
      194438    product as described, I have wasted a lot of m...
      Name: reviewText, Length: 194439, dtype: object
```

```
[13]: df.shape
```

```
[13]: (194439, 9)
```

```
[14]: review_text=df.reviewText.apply(gensim.utils.simple_preprocess)
```

```
[15]: review_text
```

```
[15]: 0      [they, look, good, and, stick, good, just, don...
      1      [these, stickers, work, like, the, review, say...
      2      [these, are, awesome, and, make, my, phone, lo...
      3      [item, arrived, in, great, time, and, was, in,...
      4      [awesome, stays, on, and, looks, great, can, b...

      ...

194434      [works, great, just, like, my, original, one, ...
194435      [great, product, great, packaging, high, quali...
194436      [this, is, great, cable, just, as, good, as, t...
194437      [really, like, it, because, it, works, well, w...
194438      [product, as, described, have, wasted, lot, of...
Name: reviewText, Length: 194439, dtype: object
```

```
[19]: review_text[0]
```

```
[19]: ['they',
      'look',
      'good',
      'and',
      'stick',
      'good',
      'just',
      'don',
      'like',
      'the',
      'rounded',
      'shape',
      'because',
      'was',
      'always',
      'bumping',
      'it',
      'and',
      'siri',
      'kept',
      'popping',
      'up',
      'and',
      'it',
      'was',
      'irritating',
      'just',
      'won',
      'buy',
      'product',
      'like',
      'this',
```

```
'again']
```

```
[20]: ###Training the Word2Vec Model
```

```
[22]: model=gensim.models.Word2Vec(  
        window=10,  
        min_count=2,  
        workers=4,  
    )
```

```
[23]: ###Build Vocabulary
```

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

```
[26]: ###Train the Word2Vec Model
```

```
[27]: model.train(review_text, total_examples=model.corpus_count, epochs=model.epochs)
```

```
[27]: (61505963, 83868975)
```

```
[28]: ##Save the Model
```

```
[29]: model.save("./word2vec-amazon-cell-accessories-reviews-short.model")
```

```
[30]: ##outpu mesure
```

```
[31]: model.wv.most_similar("bad")
```

```
[31]: [('terrible', 0.6823690533638),  
        ('shabby', 0.6281970739364624),  
        ('good', 0.5974149703979492),  
        ('horrible', 0.5920029282569885),  
        ('legit', 0.5504642128944397),  
        ('awful', 0.5471381545066833),  
        ('poor', 0.5461689233779907),  
        ('okay', 0.5404900908470154),  
        ('funny', 0.5360230803489685),  
        ('disappointing', 0.5320929288864136)]
```

```
[32]: model.wv.similarity(w1="cheap", w2="inexpensive")
```

```
[32]: 0.5480453
```

```
[33]: model.wv.similarity(w1="great", w2="good")
```

```
[33]: 0.781932
```

```
[ ]:
```