Output

Word2vec Model:

Given english-hindi dictionary

	eng	hindi
0	the	द
1	.	I
2	,	,
3	i	मैं
4	to	करने
...	...	...
142746	zx-5n	zx-5n
142747	zx-6000	zx-6000
142748	zx5000	zx5000
142749	zzzzzzzzzzzzzzz	zzzzzzzzzzzzzzz
142750	<unk>	<unk>

142751 rows × 2 columns

Hindi text after tokenisation
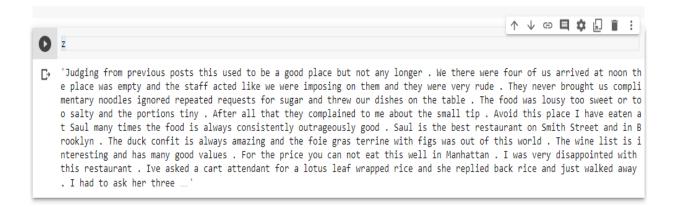
```
hin_vect[1:5]

[['इस',
  'पॉपअप',
  'बॉक्स',
  'में',
  'पासवर्ड',
  'चेंज',
  'करने',
  'लॉगइन',
  'अलर्ट्स',
  'चालू',
  'करने',
  'और',
  'मौजूदा',
  'फेसबुक',
  'सेशन',
```

Text after removing punctuation

```
m1

'Judging from previous posts this used to be a good place   but not any longer .\nWe   there were four of us   arrived at
noon  the place was empty  and the staff acted like we were imposing on them and they were very rude .\nThey never brought
us complimentary noodles   ignored repeated requests for sugar   and threw our dishes on the table .\nThe food was lousy
too sweet or too salty and the portions tiny .\nAfter all that   they complained to me about the small tip .\nAvoid this p
lace  \nI have eaten at Saul   many times   the food is always consistently   outrageously good .\nSaul is the best restau
rant on Smith Street and in Brooklyn .\nThe duck confit is always amazing and the foie gras terrine with figs was out of t
his world .\nThe wine list is interesting and has many good values .\nFor the price   you can not eat this well in Manhatt
an .\nI was very disappointed with this restaurant .\nIve asked a cart attendant for a lotus leaf wrapped rice and she rep
lied back rice and just ...'
```

Text after removing alphanumeric

```
z
```

'Judging from previous posts this used to be a good place but not any longer . We there were four of us arrived at noon th
e place was empty and the staff acted like we were imposing on them and they were very rude . They never brought us compli
mentary noodles ignored repeated requests for sugar and threw our dishes on the table . The food was lousy too sweet or to
o salty and the portions tiny . After all that they complained to me about the small tip . Avoid this place I have eaten a
t Saul many times the food is always consistently outrageously good . Saul is the best restaurant on Smith Street and in B
rooklyn . The duck confit is always amazing and the foie gras terrine with figs was out of this world . The wine list is i
nteresting and has many good values . For the price you can not eat this well in Manhattan . I was very disappointed with
this restaurant . Ive asked a cart attendant for a lotus leaf wrapped rice and she replied back rice and just walked away
. I had to ask her three …'

Initial Bingliu

	word	polarity
0	a+	positive
1	abound	positive
2	abounds	positive
3	abundance	positive
4	abundant	positive
...	...	...
6784	zaps	negative
6785	zealot	negative
6786	zealous	negative
6787	zealously	negative

Text after lowering

```
#printing the text after preprocessing
final_english
```

'judging from previous posts this used to be a good place but not any longer . we there were four of us arrived at noon th
e place was empty and the staff acted like we were imposing on them and they were very rude . they never brought us compli
mentary noodles ignored repeated requests for sugar and threw our dishes on the table . the food was lousy too sweet or to
o salty and the portions tiny . after all that they complained to me about the small tip . avoid this place i have eaten a
t saul many times the food is always consistently outrageously good . saul is the best restaurant on smith street and in b
rooklyn . the duck confit is always amazing and the foie gras terrine with figs was out of this world . the wine list is i
nteresting and has many good values . for the price you can not eat this well in manhattan . i was very disappointed with
this restaurant . ive asked a cart attendant for a lotus leaf wrapped rice and she replied back rice and just walked away
. i had to ask her three …'

After removing the stopwords

```
[ ]  #printing the text after removing the stopwords
     final_english
```

'judging from previous posts this used to be a good place but not any longer . we there were four of us arrived at noon the place was empty and the staff acted like we were imposing on them and they were very rude . they never brought us compli mentary noodles ignored repeated requests for sugar and threw our dishes on the table . the food was lousy too sweet or too salty and the portions tiny . after all that they complained to me about the small tip . avoid this place i have eaten at saul many times the food is always consistently outrageously good . saul is the best restaurant on smith street and in b rooklyn . the duck confit is always amazing and the foie gras terrine with figs was out of this world . the wine list is i nteresting and has many good values . for the price you can not eat this well in manhattan . i was very disappointed with this restaurant . ive asked a cart attendant for a lotus leaf wrapped rice and she replied back rice and just walked away . i had to ask her three ...'

After sentence tokenizing

```
#tokenizing into sentences
sents=sent_tokenize(final_english)
sents
```

```
['judging from previous posts this used to be a good place but not any longer .',
 'we there were four of us arrived at noon the place was empty and the staff acted like we were imposing on them and
 'they never brought us complimentary noodles ignored repeated requests for sugar and threw our dishes on the table
 'the food was lousy too sweet or too salty and the portions tiny .',
 'after all that they complained to me about the small tip .',
 'avoid this place i have eaten at saul many times the food is always consistently outrageously good .',
 'saul is the best restaurant on smith street and in brooklyn .',
 'the duck confit is always amazing and the foie gras terrine with figs was out of this world .',
 'the wine list is interesting and has many good values .',
 'for the price you can not eat this well in manhattan .',
 'i was very disappointed with this restaurant .',
 'ive asked a cart attendant for a lotus leaf wrapped rice and she replied back rice and just walked away .',
 'i had to ask her three times before she finally came back with the dish ive requested .',
 'food was okay nothing great
```

Tokenizing sentence wise

```
    words=word_tokenize(s)
    words_in_list.append(words)
words_in_list
```

```
 'cheescake',
 'like',
 'this',
 '.'],
['i', 'thought', 'i', 'had', 'died', 'and', 'gone', 'to', 'heaven', '.']
['great',
 'spot',
 'whether',
 'looking',
 'for',
 'a',
 'couple',
 'of',
 'drinks',
 'or',
 'quiet',
 'dinner',
 '.'],
['warm',
 'and',
 'friendly',
```

Tokenizing wordwise

```
words=word_tokenize(final_english)
words
```

```
'ot',
'the',
'half',
'price',
'sushi',
'deal',
'on',
'saturday',
'so',
```

Training the model on english file

```
actuale = gensim.models.Word2Vec(size=150, min_count=1, workers=4)#training the model for english text file
actuale.build_vocab(words_in_list)
actuale.train(words_in_list, total_examples=len(words_in_list), epochs=400)
```

```
(20801703, 29723200)
```

Vocabulory Obtained After Training Model On English Corpus

```
vocabe
```

```
'charm',
'selection',
'gigondas',
'champagne',
'may',
'packed',
'vibe',
'recommended',
'response',
'should',
'additional',
'tastes',
'freshly',
'baked',
'bread',
'veal',
```

Vocabulary after training model on hindi corpus

vocabh

```
'निकॉन',
'कैनन',
'पीछे',
'इंच',
'एलसीडी',
'सेटिंग',
'कंट्रोल',
'साधारण',
'फोर',
'पैड',
'रबर',
'आईकैप',
'आरामदायक',
'बनाया',
'एमएएच',
'रिचार्जेबल',
'लिथियम',
'एए',
```
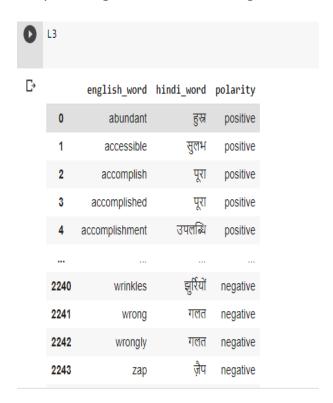
Vectors obtained after training on english


```
actuale.wv.vectors#finding out the vectors

array([[ 0.88081795,  0.27126455, -1.9439324 , ...,  0.53884006,
        -0.0176786 , -0.28982762],
       [ 1.7246131 ,  0.6969001 , -0.6973589 , ..., -0.35884902,
        -0.16412504, -0.10028429],
       [ 0.07052246, -0.9597375 , -1.4231261 , ...,  0.71135557,
         0.49529526, -0.28203246],
       ...,
       [ 0.33332187,  0.5814891 ,  1.097602  , ...,  0.5808776 ,
         2.1489587 ,  0.9616404 ],
       [ 0.8362212 , -0.00230616,  0.1864013 , ..., -0.00817429,
         1.2290337 , -0.20241956],
       [ 0.20706056, -0.11738686, -0.7304585 , ..., -1.6350286 ,
         1.3159928 , -0.786145  ]], dtype=float32)
```

Finding similar words for wine


```
actuale.wv.most_similar("wine",topn=5)#finding out the similar words for wine

/usr/local/lib/python3.6/dist-packages/gensim/matutils.py:737: FutureWarning: Conversion of the second argument of iss
  if np.issubdtype(vec.dtype, np.int):
[('extensive', 0.3002370595932007),
 ('sake', 0.2933518588542938),
 ('waste', 0.27523696422576904),
 ('dry', 0.26779428124427795),
 ('presentation', 0.2601240873336792)]
```

L3 after created with the help of BingLiu lexicon and English- hindi dictionary:

L3

	english_word	hindi_word	polarity
0	abundant	हुस्न	positive
1	accessible	सुलभ	positive
2	accomplish	पूरा	positive
3	accomplished	पूरा	positive
4	accomplishment	उपलब्धि	positive
...	...	...	...
2240	wrinkles	झुर्रियों	negative
2241	wrong	गलत	negative
2242	wrongly	गलत	negative
2243	zap	ज़ैप	negative

OUTPUT: Extended L3 which contains similar words which we got from word2vec model:

5 rows got added after training the model and getting the similar words.

L3[2244:]

	english_word	hindi_word	polarity
2244	zombie	ज़ोंबी	negative
2245	life	जीवन	positive
2246	results	परिणाम	positive
2247	use	उपयोग	positive
2248	much	ज्यादा	positive
2249	place	जगह	positive

Glove Model:

1. Cleaned text of english corpus after pre- processing (after removing punctuations, numbers and converting into lower case then tokenizing):

	sentences	clean text
0	Judging from previous posts this used to be a ...	[judging, from, previous, posts, this, used, t...
1	We , there were four of us , arrived at noon -...	[we, there, were, four, of, us, arrived, at, n...
2	They never brought us complimentary noodles , ...	[they, never, brought, us, complimentary, nood...
3	The food was lousy - too sweet or too salty an...	[the, food, was, lousy, too, sweet, or, too, s...
4	After all that , they complained to me about t...	[after, all, that, they, complained, to, me, a...
...	...	...
5286	Received this item and everything is great exc...	[received, this, item, and, everything, is, gr...
5287	To go left I have to go right and vice versa a...	[to, go, left, i, have, to, go, right, and, vi...
5288	When I attach a mouse it's fine but the touchp...	[when, i, attach, a, mouse, its, fine, but, th...
5289	I went through the settings and there isn't a ...	[i, went, through, the, settings, and, there, ...
5290	It's a great laptop, but right now it's annoyi...	[its, a, great, laptop, but, right, now, its, ...

5291 rows × 2 columns

2. Vector form after converting the processed english texts into list of lists:

[['judging', 'from', 'previous', 'posts', 'this', 'used', 'to', 'be', 'a', 'good', 'place', 'but', 'not', 'any', 'longer', ''], ['we', 'there', 'were', 'four', 'of', 'us', 'arrived', 'at', 'noon', 'the', 'place', 'was', 'empty', 'and', 'the', 'staff', 'acted', 'like', 'we', 'were', 'imposing', 'on', 'them', 'and', 'they', 'were', 'very', 'rude', ''], ['they', 'never', 'brought', 'us', 'complimentary', 'noodles', 'ignored', 'repeated', 'requests', 'for', 'sugar', 'and', 'threw', 'our', 'dishes', 'on', 'the', 'table', ''], ['the', 'food', 'was', 'lousy', 'too', 'sweet', 'or', 'too', 'salty', 'and', 'the', 'portions', 'tiny', '']...

3. Cleaned text of Hindi corpus after pre- processing (after removing punctuations, numbers and English alphabets, then tokenizing):

	sentences	clean text
0	फेसबुक का सिक्योरिटी चैकअप फीचर पॉपअप की तरह य...	[फेसबुक, का, सिक्योरिटी, चैकअप, फीचर, पॉपअप, क...
1	इस पॉपअप बॉक्स में पासवर्ड चेंज करने , लॉगिन अ...	[इस, पॉपअप, बॉक्स, में, पासवर्ड, चेंज, करने, ल...
2	अब नए फीचर से इस तरह की गलतियां नहीं होंगी और ...	[अब, नए, फीचर, से, इस, तरह, की, गलतियां, नहीं,...
3	Twitter periscope App एंड्रॉयड 4.4.4 किटकैट या...	[एंड्रॉयड, किटकैट, या, उससे, ऊपर, के, वर्जन, प...
4	पेरिस्कोप एप में रिज्यूम नोटिफिकेशन फीचर दिया ...	[पेरिस्कोप, एप, में, रिज्यूम, नोटिफिकेशन, फीचर...
...	...	...
5412	तीसरा जो है वो इसके आगे के दशक का लगता है । ...	[तीसरा, जो, है, वो, इसके, आगे, के, दशक, का, लग...
5413	यह वो बातें हैं जो फिल्म को रोचक बनाने के साथ ...	[यह, वो, बातें, हैं, जो, फिल्म, को, रोचक, बनान...
5414	वैसे इस बात के अनेक उदाहरण हमारे सामने हैं कि ...	[वैसे, इस, बात, के, अनेक, उदाहरण, हमारे, सामने...
5415	शेखर कपूर ही एक अपवाद रहे हैं जिन्होंने हॉलीवु...	[शेखर, कपूर, ही, एक, अपवाद, रहे, हैं, जिन्होंन...
5416	उनके अलावा कोई भी इस रेखा को पार नहीं कर पाया ...	[उनके, अलावा, कोई, भी, इस, रेखा, को, पार, नहीं...

5417 rows × 2 columns

4. Vector after converting the processed hindi text into list of lists:

[['फेसबुक', 'का', 'सिक्योरिटी', 'चैकअप', 'फीचर', 'पॉपअप', 'की', 'तरह', 'यूजर्स', 'को', 'दिखाइ', 'देगा', '।'], ['इस', 'पॉपअप', 'बॉक्स', 'में', 'पासवर्ड', 'चेंज', 'करने', 'लॉगिन', 'अलर्ट्स', 'चालू', 'करने', 'और', 'मौजूदा', 'फेसबुक', 'सेशन', 'को', 'पूरी', 'तरह', 'से', 'बंद', 'करने', 'के', 'ऑप्शन', 'मिलेंगे', '।'],

5. Embeddings of English words:

[array([-0.19981594, 0.48287539, 0.11798714, 0.36544131]), array([0.68184564, 0.07201014, -0.89335174, -0.22678978]), array([0.19645565, 0.58520876, 0.04642479, 0.40520817]), array([-0.18951875, 0.28158477, 0.11030827, 0.26409446]), array([0.77171873, -0.66855201, -1.508529 , -0.85797281]), array([0.26828222, 0.21536538, -0.50294374, -0.93340784]), array([0.54868583, -0.1805312 , -1.75732618, -1.56283972]),

6. Embeddings of Hindi words:

[array([-0.06742127, -0.25138595, 0.16664987, -0.17922376]), array([-1.13705136, 0.20126498, -0.33905654, 1.53341795]), array([0.28687086, -0.40743695, 0.28142187, 0.02354826]), array([0.39833934, -0.25640939, 0.32379015, -0.04912226]), array([-0.32062469, 0.38278636, -0.18627394, 0.55760482]), array([-0.09336768, -0.12788815, 0.15069697, -0.07715594]), array([-1.76049787, 1.21533006, -0.06672176, 0.48685255]), array([-1.22077257, 0.66348149, -0.40270695, 0.39011677]),

7. Similar words :

```python
print(gloveE.most_similar('food'))
print(gloveH.most_similar('तस्वीर'))
```

[('machine', 0.9986149930116509), ('ready', 0.9896818466817 18), ('cool', 0.9871403536066438), ('dinner', 0.9861013084673559)]
[('कहते', 0.9906361812488317), ('मेहनत', 0.9863545388867683), ('दिए', 0.9847142417368085), ('कालिंग', 0.9827822091449058)]