NLP Task - Feb 17

20BCE2685

In [10]:

```
from sklearn.feature_extraction.text import CountVectorizer
vect = CountVectorizer(binary = True)
```

In [11]:

```
corpus = ["A new-fangled tea drinker (black, two sugars, no milk please!)","I recent
```

In [12]:

```
vect.fit(corpus)
```

Out[12]:

```
CountVectorizer(binary=True)
```

In [13]:

```
vocab = vect.vocabulary_
```

In [14]:

```
for key in sorted(vocab.keys()):
    print("{}:{}".format(key,vocab[key]))
```

```
about:0
aforementioned:1
and:2
as:3
assam:4
banks:5
belt:6
bewildered:7
black:8
brahmaputra:9
but:10
class:11
could:12
drinker:13
fangled:14
fertile:15
found:16
grappling:17
india:18
its:19
later:20
learn:21
least:22
lies:23
like:24
little:25
male:26
me:27
mighty:28
milk:29
mission:30
more:31
much:32
my:33
myself:34
named:35
```

```
new:36
no:37
of:38
on:39
only:40
phrases:41
please:42
possibly:43
premier:44
producing:45
recently:46
region:47
river:48
say:49
south:50
strange:51
sugars:52
tea:53
teas:54
terminologies:55
that:56
the:57
those:58
three:59
to:60
trip:61
two:62
upper:63
various:64
very:65
was:66
with:67
words:68
world:69
```

In [16]:

```
print(vect.transform(["I was on a trip to India's premier tea producing belt of Uppe
```

```
[[0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
  0 0 1 1 0 0 0 0 1 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 1 0 1 0 0 1 0 0 0]]
```

In [17]:

```
from sklearn.metrics.pairwise import cosine_similarity
similarity = cosine_similarity(vect.transform(["I was on a trip to India's premier t
```

In [18]:

```
print(similarity)
```

```
[[0.07216878]]
```

In [ ]:

In [ ]:

In [ ]: