

# ML LAB Experiment 8 - Text Summarization using NLP Based Approaches

- Name - Pravesh P. Ganwani
- T.E. I.T.
- Batch B
- Roll No. 2018140021

## What is Text Summarization?

Text Summarization involves condensing a piece of text into a shorter version, reducing the size of the original text while preserving key information and the meaning of the content. Since manual text synthesis is a long and generally laborious task, task automation is gaining in popularity

## Types of Approaches to Summarize Text -

1. **Extractive Approach** - The Extractive approach takes sentences directly from the document according to a scoring function to form a cohesive summary. This method works by identifying the important sections of the text cropping and assembling parts of the content to produce a condensed version.
2. **Abstractive Approach** - The Abstraction approach aims to produce a summary by interpreting the text using advanced natural language techniques to generate a new, shorter text – parts of which may not appear in the original document, which conveys the most information.

**Dataset** - Amazon Fine Food Reviews Dataset

**Dataset Description** - This dataset consists of reviews of fine foods from amazon. The data span a period of more than 10 years, including all ~500,000 reviews up to October 2012. Reviews include product and user information, ratings, and a plain text review. It also includes reviews from all other Amazon categories.

In [ ]: *# Importing all required libraries*

```
import pandas as pd
import numpy as np
import nltk # Natural Language Toolkit
nltk.download('punkt') # Sentence Tokenizer
nltk.download('stopwords') # Stop Words
import re
from nltk.corpus import stopwords
```

[nltk\_data] Downloading package punkt to /root/nltk\_data...

[nltk\_data] Package punkt is already up-to-date!

[nltk\_data] Downloading package stopwords to /root/nltk\_data...

[nltk\_data] Package stopwords is already up-to-date!

```
In [ ]: # Reading CSV

df = pd.read_csv("/content/drive/MyDrive/ML Experiment 8/amzn_reviews.csv")
df.head()
```

Out[ ]:

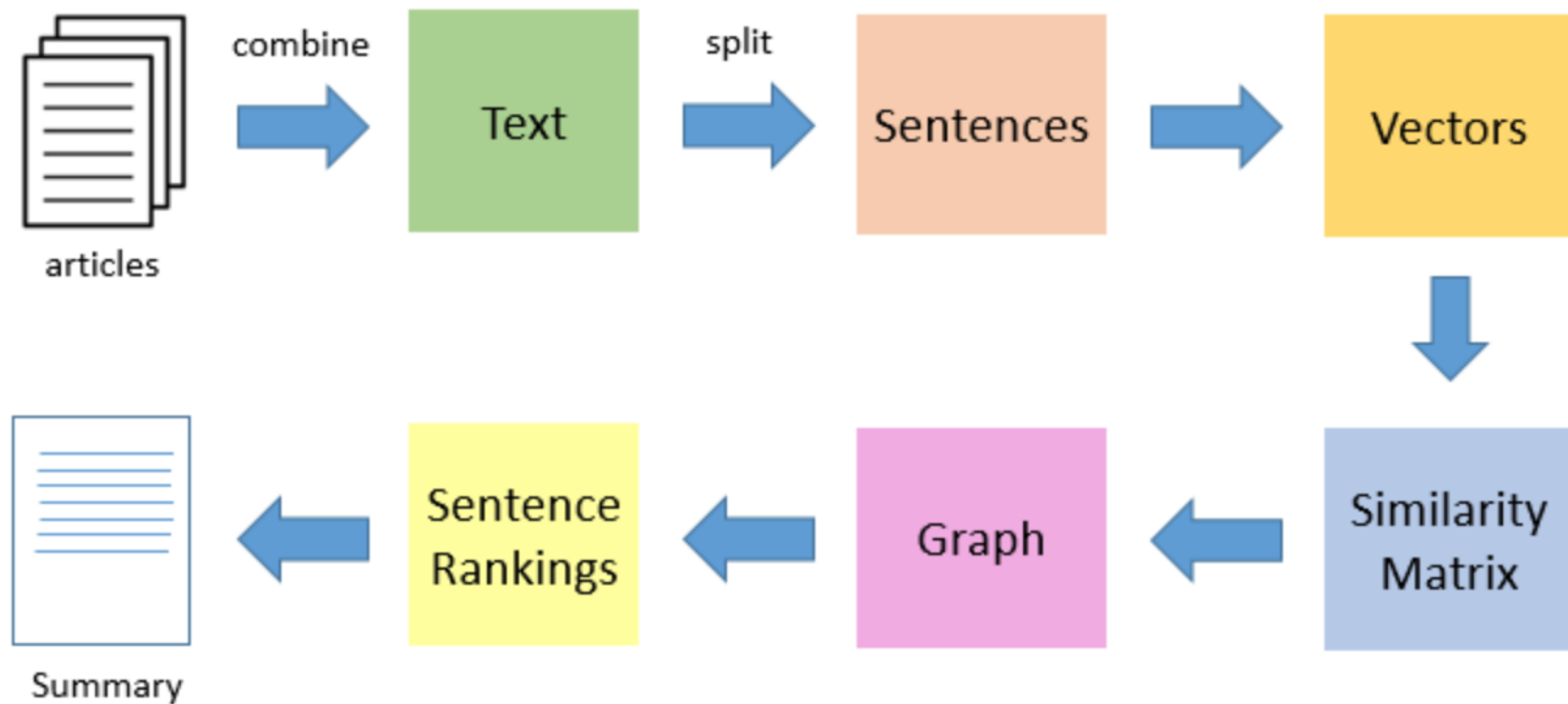
	Id	ProductId	UserId	ProfileName	HelpfulnessNumerator	HelpfulnessDenominator	Score	Time	Summary
0	1	B001E4KFG0	A3SGXH7AUHU8GW	delmartian	1	1	5	1303862400	Good Quality Dog Food
1	2	B00813GRG4	A1D87F6ZCVE5NK	dll pa	0	0	1	1346976000	Not as Advertised
2	3	B000LQOCH0	ABXLMWJIXXAIN	Natalia Corres "Natalia Corres"	1	1	4	1219017600	"Delight" says it all
3	4	B000UA0QIQ	A395BORC6FGVXV	Karl	3	3	2	1307923200	Cough Medicine
4	5	B006K2ZZ7K	A1UQRSCLF8GW1T	Michael D. Bigham "M. Wassir"	0	0	5	1350777600	Great taffy

```
In [ ]: df['Text'][1] # Full article text
```

```
Out[ ]: 'Product arrived labeled as Jumbo Salted Peanuts...the peanuts were actually small sized unsalted. Not sure if this was an error or if the vendor intended to represent the product as "Jumbo".'
```

## Extractive Approach to Summarize Text

Extractive Text summarization is the processing of creating a summary from text using phrases and sentences from the document and stacking them up to create a summary.



### Sentence Splitting using Sentence Tokenizer

```
In [8]: from nltk.tokenize import sent_tokenize

sentences = []
for s in df['Text']:
    sentences.extend(sent_tokenize(s)) # Appending split sentences to a list
```

In [20]: sentences

```
Out[20]: ['I have bought several of the Vitality canned dog food products and have found them all to be of good quality.',
'The product looks more like a stew than a processed meat and it smells better.',
'My Labrador is finicky and she appreciates this product better than most.',
'Product arrived labeled as Jumbo Salted Peanuts...the peanuts were actually small sized unsalted.',
'Not sure if this was an error or if the vendor intended to represent the product as "Jumbo".',
'This is a confection that has been around a few centuries.',
'It is a light, pillowy citrus gelatin with nuts - in this case Filberts.',
'And it is cut into tiny squares and then liberally coated with powdered sugar.',
'And it is a tiny mouthful of heaven.',
'Not too chewy, and very flavorful.',
'I highly recommend this yummy treat.',
'If you are familiar with the story of C.S.',
'Lewis\' "The Lion, The Witch, and The Wardrobe" - this is the treat that seduces Edmund into selling out his Brother and Sisters to the Witch.',
'If you are looking for the secret ingredient in Robitussin I believe I have found it.',
'I got this in addition to the Root Beer Extract I ordered (which was good) and made some cherry soda.',
'The flavor is very medicinal.',
'Great taffy at a great price.',
'There was a wide assortment of yummy taffy.',
'Delivery was very quick.',
'If your a taffy lover, this is a deal.',
'I got a wild hair for taffy and ordered this five pound bag.',
'The taffy was all very enjoyable with many flavors: watermelon, root beer, melon, peppermint, grape, etc.',
'My only complaint is there was a bit too much red/black licorice-flavored pieces (just not my particular favorites).',
'Between me, my kids, and my husband, this lasted only two weeks!',
'I would recommend this brand of taffy -- it was a delightful treat.',
'This saltwater taffy had great flavors and was very soft and chewy.',
'Each candy was individually wrapped well.',
'None of the candies were stuck together, which did happen in the expensive version, Fralinger's.',
'Would highly recommend this candy!',
'I served it at a beach-themed party and everyone loved it!',
'This taffy is so good.',
'It is very soft and chewy.',
'The flavors are amazing.',
'I would definitely recommend you buying it.',
'Very satisfying!',
'!',
'Right now I'm mostly just sprouting this so my cats can eat the grass.',
'They love it.',
'I rotate it around with Wheatgrass and Rye too',
'This is a very healthy dog food.',
```

'Good for their digestion.',  
'Also good for small puppies.',  
'My dog eats her required amount at every feeding.',  
"I don't know if it's the cactus or the tequila or just the unique combination of ingredients, but the flavor of this hot sauce makes it one of a kind!",  
'We picked up a bottle once on a trip we were on and brought it back home with us and were totally blown away!',  
"When we realized that we simply couldn't find it anywhere in our city we were bummed.<br /><br />Now, because of the magic of the internet, we have a case of the sauce and are ecstatic because of it.<br /><br />If you love hot sauce..I mean really love hot sauce, but don't want a sauce that tastelessly burns your throat, grab a bottle of Tequila Picante Gourmet de Incan.",  
'Just realize that once you taste it, you will never want to use any other sauce.<br /><br />Thank you for the personal, incredible service!',  
"One of my boys needed to lose some weight and the other didn't.",  
'I put this food on the floor for the chubby guy, and the protein-rich, no by-product food up higher where only my skinny boy can jump.',  
'The higher food sits going stale.',  
'They both really go for this food.',  
'And my chubby boy has been losing about an ounce a week.',  
'My cats have been happily eating Felidae Platinum for more than two years.',  
'I just got a new bag and the shape of the food is different.',  
'They tried the new food when I first put it in their bowls and now the bowls sit full and the kitties will not touch the food.',  
"I've noticed similar reviews related to formula changes in the past.",  
'Unfortunately, I now need to find a new food that my cats will eat.',  
'good flavor!',  
'these came securely packed... they were fresh and delicious!',  
'i love these Twizzlers!',  
'The Strawberry Twizzlers are my guilty pleasure - yummy.',  
'Six pounds will be around for a while with my son and I.',  
'My daughter loves twizzlers and this shipment of six pounds really hit the spot.',  
"It's exactly what you would expect...six packages of strawberry twizzlers.",  
'I love eating them and they are good for watching TV and looking at movies!',  
'It is not too sweet.',  
'I like to transfer them to a zip lock baggie so they stay fresh so I can take my time eating them.',  
'I am very satisfied with my Twizzler purchase.',  
'I shared these with others and we have all enjoyed them.',  
'I will definitely be ordering more.',  
'Twizzlers, Strawberry my childhood favorite candy, made in Lancaster Pennsylvania by Y & S Candies, Inc. one of the oldest confectionery Firms in the United States, now a Subsidiary of the Hershey Company, the Company was established in 1845 as Young and Smylie, they also make Apple Licorice Twists, Green Color and Blue Raspberry Licorice Twists, I like them all<br /><br />I keep it in a dry cool place because is not recommended it to put it in the fridge.',



'According to the Guinness Book of Records, the longest Licorice Twist ever made measured 1.200 Feet (370 M) and weighted 100 Pounds (45 Kg) and was made by Y & S Candies, Inc.',  
'This Record-Breaking Twist became a Guinness World Record on July 19, 1998.',  
'This Product is Kosher!'  
'Thank You',  
'Candy was delivered very fast and was purchased at a reasonable price.',  
'I was home bound and unable to get to a store so this was perfect for me.',  
'My husband is a Twizzlers addict.',  
"We've bought these many times from Amazon because we're government employees living overseas and can't get them in the country we are assigned to.",  
"They've always been fresh and tasty, packed well and arrive in a timely manner.",  
'I bought these for my husband who is currently overseas.',  
'He loves these, and apparently his staff likes them also.<br />There are generous amounts of Twizzlers in each 16-ounce bag, and this was well worth the price.',  
'<a href="http://www.amazon.com/gp/product/B001GVISJM">Twizzlers, Strawberry, 16-Ounce Bags (Pack of 6)</a>',  
>',  
"I can remember buying this candy as a kid and the quality hasn't dropped in all these years.",  
"Still a superb product you won't be disappointed with.",  
'I love this candy.',  
'After weight watchers I had to cut back but still have a craving for it.',  
'I have lived out of the US for over 7 yrs now, and I so miss my Twizzlers!!',  
'When I go back to visit or someone visits me, I always stock up.',  
"All I can say is YUM!<br />Sell these in Mexico and you will have a faithful buyer, more often than I'm able to buy them right now.",  
'Product received is as advertised.<br /><br /><a href="http://www.amazon.com/gp/product/B001GVISJM">Twizzlers, Strawberry, 16-Ounce Bags (Pack of 6)</a>',  
'The candy is just red , No flavor .',  
'Just plan and chewy .',  
'I would never buy them again',  
'I was so glad Amazon carried these batteries.',  
'I have a hard time finding them elsewhere because they are such a unique size.',  
'I need them for my garage door opener.<br />Great deal for the price.',  
"I got this for my Mum who is not diabetic but needs to watch her sugar intake, and my father who simply chooses to limit unnecessary sugar intake - she's the one with the sweet tooth - they both LOVED these toffees, you would never guess that they're sugar-free and it's so great that you can eat them pretty much guilt free!",  
"i was so impressed that i've ordered some for myself (w dark chocolate) to take to the office so i'll eat them instead of snacking on sugary sweets.<br />These are just EXCELLENT!",  
"I don't know if it's the cactus or the tequila or just the unique combination of ingredients, but the flavor of this hot sauce makes it one of a kind!",  
'We picked up a bottle once on a trip we were on and brought it back home with us and were totally blown away!',  
"When we realized that we simply couldn't find it anywhere in our city we were bummed.<br /><br />Now, because

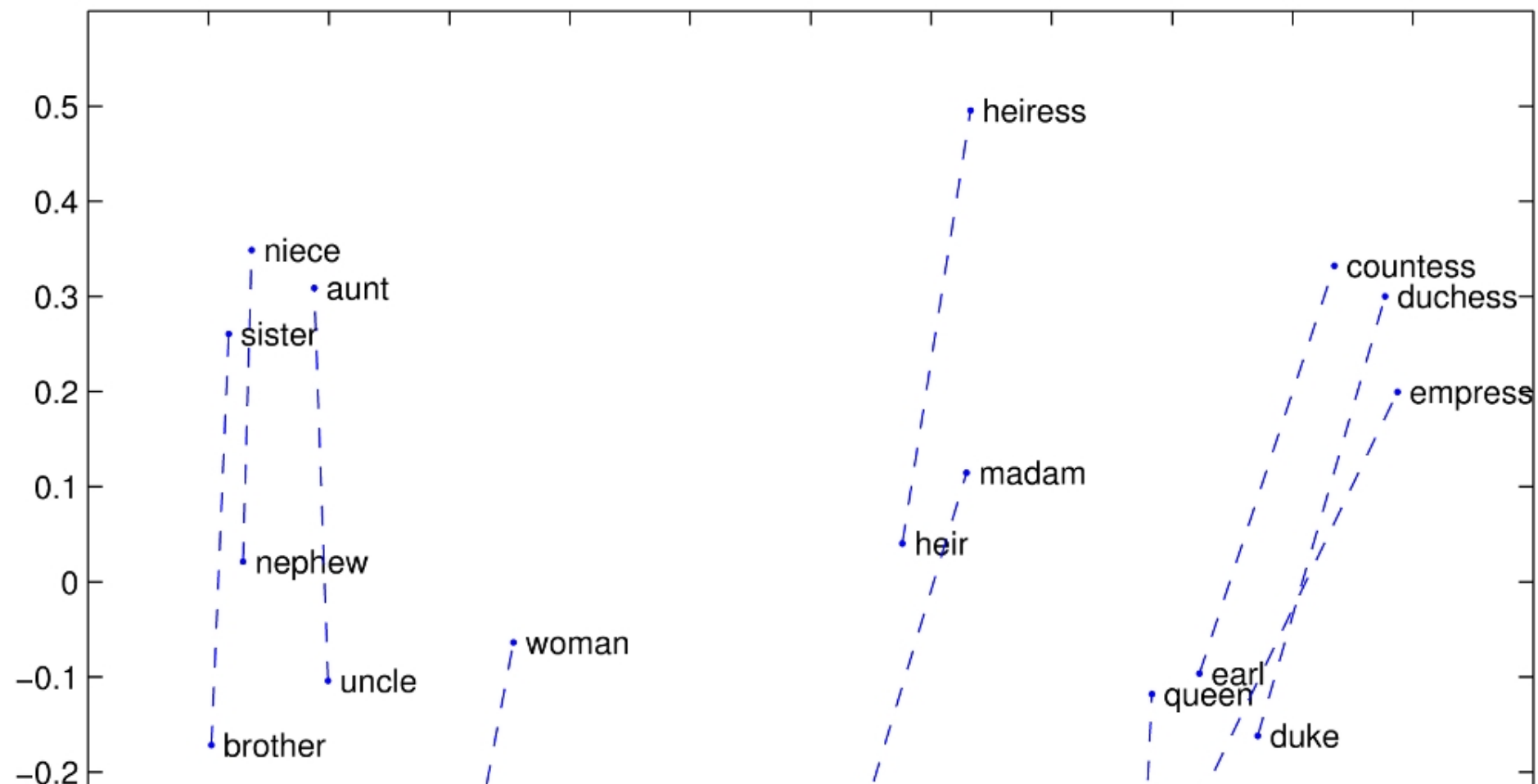
se of the magic of the internet, we have a case of the sauce and are ecstatic because of it.<br /><br />If you love hot sauce..I mean really love hot sauce, but don't want a sauce that tastelessly burns your throat, grab a bottle of Tequila Picante Gourmet de Inclan.",  
'Just realize that once you taste it, you will never want to use any other sauce.<br /><br />Thank you for the personal, incredible service!']

In [9]: sentences[:5]

Out[9]: ['I have bought several of the Vitality canned dog food products and have found them all to be of good quality.',  
'The product looks more like a stew than a processed meat and it smells better.',  
'My Labrador is finicky and she appreciates this product better than most.',  
'Product arrived labeled as Jumbo Salted Peanuts...the peanuts were actually small sized unsalted.',  
'Not sure if this was an error or if the vendor intended to represent the product as "Jumbo".']

**Using GLOVE for Word Vector Representation.**

Glove is an unsupervised learning algorithm developed by Stanford for generating word embeddings by aggregating global word-word co-occurrence matrix from a corpus.



```
In [10]: word_embeddings = {}  
ctr = 0  
f = open('/content/drive/MyDrive/ML Experiment 8/glove.6B.100d.txt', encoding='utf-8')  
for line in f:  
    values = line.split()  
    if(ctr < 5):  
        print(values)  
        ctr+=1  
    word = values[0]  
    coefs = np.asarray(values[1:], dtype='float32')  
    word_embeddings[word] = coefs  
f.close()
```

['the', '-0.038194', '-0.24487', '0.72812', '-0.39961', '0.083172', '0.043953', '-0.39141', '0.3344', '-0.57545', '0.087459', '0.28787', '-0.06731', '0.30906', '-0.26384', '-0.13231', '-0.20757', '0.33395', '-0.33848', '-0.31743', '-0.48336', '0.1464', '-0.37304', '0.34577', '0.052041', '0.44946', '-0.46971', '0.02628', '-0.54155', '-0.15518', '-0.14107', '-0.039722', '0.28277', '0.14393', '0.23464', '-0.31021', '0.086173', '0.20397', '0.52624', '0.17164', '-0.082378', '-0.71787', '-0.41531', '0.20335', '-0.12763', '0.41367', '0.55187', '0.57908', '-0.33477', '-0.36559', '-0.54857', '-0.062892', '0.26584', '0.30205', '0.99775', '-0.80481', '-3.0243', '0.01254', '-0.36942', '2.2167', '0.72201', '-0.24978', '0.92136', '0.034514', '0.46745', '1.1079', '-0.19358', '-0.074575', '0.23353', '-0.052062', '-0.22044', '0.057162', '-0.15806', '-0.30798', '-0.41625', '0.37972', '0.15006', '-0.53212', '-0.2055', '-1.2526', '0.071624', '0.70565', '0.49744', '-0.42063', '0.26148', '-1.538', '-0.30223', '-0.073438', '-0.28312', '0.37104', '-0.25217', '0.016215', '-0.017099', '-0.38984', '0.87424', '-0.72569', '-0.51058', '-0.52028', '-0.1459', '0.8278', '0.27062']

['', '-0.10767', '0.11053', '0.59812', '-0.54361', '0.67396', '0.10663', '0.038867', '0.35481', '0.06351', '-0.094189', '0.15786', '-0.81665', '0.14172', '0.21939', '0.58505', '-0.52158', '0.22783', '-0.16642', '-0.68228', '0.3587', '0.42568', '0.19021', '0.91963', '0.57555', '0.46185', '0.42363', '-0.095399', '-0.42749', '-0.16567', '-0.056842', '-0.29595', '0.26037', '-0.26606', '-0.070404', '-0.27662', '0.15821', '0.69825', '0.43081', '0.27952', '-0.45437', '-0.33801', '-0.58184', '0.22364', '-0.5778', '-0.26862', '-0.20425', '0.56394', '-0.58524', '-0.14365', '-0.64218', '0.0054697', '-0.35248', '0.16162', '1.1796', '-0.47674', '-2.7553', '-0.1321', '-0.047729', '1.0655', '1.1034', '-0.2208', '0.18669', '0.13177', '0.15117', '0.7131', '-0.35215', '0.91348', '0.61783', '0.70992', '0.23955', '-0.14571', '-0.37859', '-0.045959', '-0.47368', '0.2385', '0.20536', '-0.18996', '0.32507', '-1.1112', '-0.36341', '0.98679', '-0.084776', '-0.54008', '0.11726', '-1.0194', '-0.24424', '0.12771', '0.013884', '0.080374', '-0.35414', '0.34951', '-0.7226', '0.37549', '0.4441', '-0.99059', '0.61214', '-0.35111', '-0.83155', '0.45293', '0.082577']

['.', '-0.33979', '0.20941', '0.46348', '-0.64792', '-0.38377', '0.038034', '0.17127', '0.15978', '0.46619', '-0.019169', '0.41479', '-0.34349', '0.26872', '0.04464', '0.42131', '-0.41032', '0.15459', '0.022239', '-0.64653', '0.25256', '0.043136', '-0.19445', '0.46516', '0.45651', '0.68588', '0.091295', '0.21875', '-0.70351', '0.16785', '-0.35079', '-0.12634', '0.66384', '-0.2582', '0.036542', '-0.13605', '0.40253', '0.14289', '0.38132', '-0.12283', '-0.45886', '-0.25282', '-0.30432', '-0.11215', '-0.26182', '-0.22482', '-0.44554', '0.2991', '-0.85612', '-0.14503', '-0.49086', '0.0082973', '-0.17491', '0.27524', '1.4401', '-0.21239', '-2.8435', '-0.27958', '-0.45722', '1.6386', '0.78808', '-0.55262', '0.65', '0.086426', '0.39012', '1.0632', '-0.35379', '0.48328', '0.346', '0.84174', '0.098707', '-0.24213', '-0.27053', '0.045287', '-0.40147', '0.11395', '0.0062226', '0.036673', '0.018518', '-1.0213', '-0.20806', '0.64072', '-0.068763', '-0.58635', '0.33476', '-1.1432', '-0.1148', '-0.25091', '-0.45907', '-0.096819', '-0.17946', '-0.063351', '-0.67412', '-0.068895', '0.53604', '-0.87773', '0.31802', '-0.39242', '-0.23394', '0.47298', '-0.028803']

['of', '-0.1529', '-0.24279', '0.89837', '0.16996', '0.53516', '0.48784', '-0.58826', '-0.17982', '-1.3581', '0.42541', '0.15377', '0.24215', '0.13474', '0.41193', '0.67043', '-0.56418', '0.42985', '-0.012183', '-0.11677', '0.31781', '0.054177', '-0.054273', '0.35516', '-0.30241', '0.31434', '-0.33846', '0.71715', '-0.26855', '-0.15837', '-0.47467', '0.051581', '-0.33252', '0.15003', '-0.1299', '-0.54617', '-0.37843', '0.64261', '0.82187', '-0.080006', '0.078479', '-0.96976', '-0.57741', '0.56491', '-0.39873', '-0.057099', '0.19743', '0.065706', '-0.48092', '-0.20125', '-0.40834', '0.39456', '-0.02642', '-0.11838', '1.012', '-0.53171', '-2.7474', '-0.042981', '-0.74849', '1.7574', '0.59085', '0.04885', '0.78267', '0.38497', '0.42097', '0.67882', '0.10337', '0.6328', '-0.026595', '0.58647', '-0.44332', '0.33057', '-0.12022', '-0.55645', '0.073611', '0.20915', '0.43395', '-0.012761', '0.089874', '-1.7991', '0.084808', '0.77112', '0.63105', '-0.90685', '0.60326', '-1.7515', '0.18596', '-0.50687', '-0.70203', '0.66578', '-0.81304', '0.18712', '-0.018488', '-0.26757', '0.727',

```
'-0.59363', '-0.34839', '-0.56094', '-0.591', '1.0039', '0.20664']
['to', '-0.1897', '0.050024', '0.19084', '-0.049184', '-0.089737', '0.21006', '-0.54952', '0.098377', '-0.201
35', '0.34241', '-0.092677', '0.161', '-0.13268', '-0.2816', '0.18737', '-0.42959', '0.96039', '0.13972', '-
1.0781', '0.40518', '0.50539', '-0.55064', '0.4844', '0.38044', '-0.0029055', '-0.34942', '-0.099696', '-0.78
368', '1.0363', '-0.2314', '-0.47121', '0.57126', '-0.21454', '0.35958', '-0.48319', '1.0875', '0.28524', '0.
12447', '-0.039248', '-0.076732', '-0.76343', '-0.32409', '-0.5749', '-1.0893', '-0.41811', '0.4512', '0.1211
2', '-0.51367', '-0.13349', '-1.1378', '-0.28768', '0.16774', '0.55804', '1.5387', '0.018859', '-2.9721', '-
0.24216', '-0.92495', '2.1992', '0.28234', '-0.3478', '0.51621', '-0.43387', '0.36852', '0.74573', '0.07210
2', '0.27931', '0.92569', '-0.050336', '-0.85856', '-0.1358', '-0.92551', '-0.33991', '-1.0394', '-0.067203',
'-0.21379', '-0.4769', '0.21377', '-0.84008', '0.052536', '0.59298', '0.29604', '-0.67644', '0.13916', '-1.55
04', '-0.20765', '0.7222', '0.52056', '-0.076221', '-0.15194', '-0.13134', '0.058617', '-0.31869', '-0.6141
9', '-0.62393', '-0.41548', '-0.038175', '-0.39804', '0.47647', '-0.15983']
```

## Text Preprocessing

```
In [11]: clean_sentences = pd.Series(sentences).str.replace("[^a-zA-Z]", " ") # Characters to removed
clean_sentences = [s.lower() for s in clean_sentences]
stop_words = stopwords.words('english')

# Function to remove stopwords
def remove_stopwords(sen):
    sen_new = " ".join([i for i in sen if i not in stop_words])
    return sen_new

clean_sentences = [remove_stopwords(r.split()) for r in clean_sentences]
```

```
In [12]: clean_sentences[:5]
```

```
Out[12]: ['bought several vitality canned dog food products found good quality',
'product looks like stew processed meat smells better',
'labrador finicky appreciates product better',
'product arrived labeled jumbo salted peanuts peanuts actually small sized unsalted',
'sure error vendor intended represent product jumbo']
```

## Creating Vectors for All Sentences

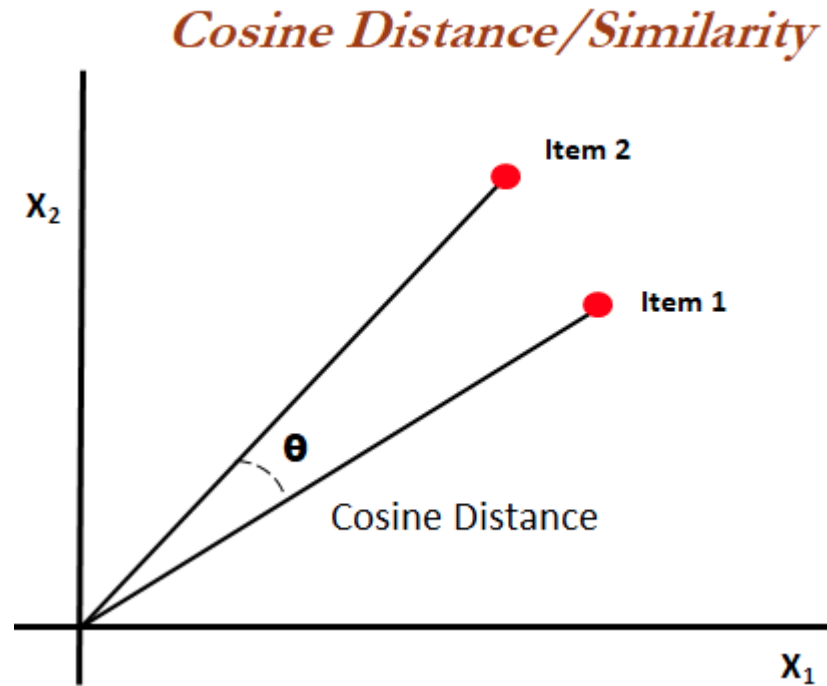
```
In [13]: sentence_vectors = []
         for i in clean_sentences:
             if len(i) != 0:
                 v = sum([word_embeddings.get(w, np.zeros((100,))) for w in i.split()])/(len(i.split())+0.001) # Adding ea
                 ch word vector for getting the summarized word vector for every sentence
             else:
                 v = np.zeros((100,)) # If no word vector found, add a zero numpy array
         sentence_vectors.append(v)
```

```
In [14]: sentence_vectors[:1]
```

```
Out[14]: [array([ 5.78533821e-02,  3.04994673e-01,  1.74281657e-01, -2.53920406e-01,
                  -3.82951871e-02, -1.37533352e-01,  2.97570705e-01,  3.20607945e-02,
                  -1.14926342e-02, -1.21462755e-01, -2.12353438e-01, -1.34225190e-01,
                   3.03125679e-01,  1.63664520e-01,  6.37719110e-02,  8.64169523e-02,
                   4.47919995e-01,  3.33842486e-02, -8.46176520e-02,  2.98351556e-01,
                   6.27958924e-02,  1.69356063e-01,  1.76763326e-01, -2.39134610e-01,
                  -7.00865090e-02,  3.49170059e-01, -1.51070878e-01, -2.76036412e-01,
                  -5.58947742e-01,  1.21637419e-01,  1.77197993e-01,  3.98444414e-01,
                  -2.68207550e-01,  3.43869589e-02,  2.12165475e-01,  3.94834548e-01,
                   9.71806794e-02,  2.25867987e-01, -2.22202748e-01, -4.27532971e-01,
                   7.08379643e-03, -5.76266408e-01, -1.08159617e-01, -3.99665415e-01,
                   1.85321078e-01,  2.04323575e-01, -2.82532632e-01, -1.80586845e-01,
                   4.09683920e-02, -7.14760482e-01,  8.74909461e-02,  4.86821271e-02,
                  -7.39393085e-02,  6.86290741e-01, -5.20413890e-02, -1.48287582e+00,
                  -3.24764580e-01, -4.67480905e-02,  1.33268285e+00,  1.19417999e-02,
                   1.64708734e-01,  7.74119020e-01, -1.35220766e-01,  2.58705229e-01,
                   5.22770882e-01,  1.39823579e-03,  4.13056672e-01, -1.14716925e-01,
                   5.62804639e-01, -2.65300244e-01, -7.73192644e-02,  1.63192868e-01,
                   2.88577110e-01,  9.75525379e-02,  2.04384044e-01,  4.51246083e-01,
                  -2.66747147e-01, -2.32935715e-02, -6.71585202e-01,  1.17823213e-01,
                   5.65074503e-01,  5.81532829e-02, -4.80116010e-01,  3.63200694e-01,
                  -1.04253471e+00, -1.61913797e-01,  1.00888088e-01, -2.98101250e-02,
                  -4.65012461e-01, -1.76576123e-01,  1.13736458e-01, -2.05094486e-01,
                   4.64616567e-02, -1.18684016e-01, -6.65627956e-01, -7.72717670e-02,
                  -3.91822994e-01, -3.46366763e-01,  7.12939382e-01,  3.88058186e-01],
                  dtype=float32)]
```



## Finding Similarities Between Sentences using Cosine Similarity



```
In [15]: sim_mat = np.zeros([len(sentences), len(sentences)]) # Empty matrix for storing similarity values

from sklearn.metrics.pairwise import cosine_similarity

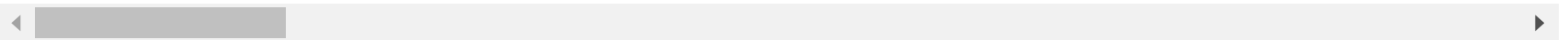
for i in range(len(sentences)):
    for j in range(len(sentences)):
        if i != j:
            sim_mat[i][j] = cosine_similarity(sentence_vectors[i].reshape(1,100), sentence_vectors[j].reshape(1,100))
            # Calculating similarity values for every sentence
```

```
In [16]: df2 = pd.DataFrame(sim_mat)
df2
```

Out[16]:

	0	1	2	3	4	5	6	7	8	9	10	11	12
0	0.000000	0.865825	0.708841	0.726431	0.760700	0.567876	0.618709	0.590632	0.435032	0.221236	0.618635	0.588630	0.556846
1	0.865825	0.000000	0.754659	0.723209	0.723266	0.560396	0.682901	0.640980	0.549292	0.449460	0.669550	0.537338	0.508773
2	0.708841	0.754659	0.000000	0.488853	0.633587	0.341855	0.421730	0.398665	0.338907	0.367956	0.624643	0.374008	0.349318
3	0.726431	0.723209	0.488853	0.000000	0.651178	0.492920	0.669380	0.736221	0.491901	0.329349	0.496180	0.466459	0.396145
4	0.760700	0.723266	0.633587	0.651178	0.000000	0.474438	0.453477	0.457732	0.391069	0.062324	0.551231	0.617773	0.426947
...	...	...	...	...	...	...	...	...	...	...	...	...	...
98	0.821443	0.801838	0.613012	0.733386	0.674644	0.558038	0.639741	0.666315	0.565224	0.362854	0.664028	0.643360	0.625274
99	0.816918	0.863213	0.703512	0.699636	0.717307	0.593550	0.723820	0.648394	0.574425	0.413469	0.643994	0.626602	0.503562
100	0.758725	0.674772	0.457192	0.604803	0.703700	0.644312	0.460573	0.475456	0.518399	0.012863	0.424999	0.632371	0.575350
101	0.761858	0.796485	0.611304	0.677120	0.670906	0.584309	0.632823	0.612234	0.575502	0.316780	0.609804	0.670932	0.614274
102	0.782205	0.748728	0.668649	0.607913	0.748356	0.530931	0.534164	0.488708	0.448268	0.196976	0.617866	0.685059	0.584298

103 rows × 103 columns



## Converting Similarity Matrix to a Network Graph using NetworkX

Nodes of this graph will represent the sentences and the edges will represent the similarity scores between the sentences.

```
In [17]: import networkx as nx
import matplotlib.pyplot as plt

nx_graph = nx.from_numpy_array(sim_mat)
plt.figure(3,figsize=(20,20))
nx.draw_random(nx_graph, with_labels=True)
```

Output hidden; open in <https://colab.research.google.com> to view.

## Summary Extraction

Extraction of summaries from text based on their rankings for summary generation.

```
In [18]: scores = nx.pagerank(nx_graph)
scores
```

```
Out[18]: {0: 0.011363842644330402,  
1: 0.011140570518279798,  
2: 0.009217590489292707,  
3: 0.009659130193868583,  
4: 0.01033199595784022,  
5: 0.00853471941472121,  
6: 0.008351750313671847,  
7: 0.008346515223468891,  
8: 0.008123595327268229,  
9: 0.0047202546527219495,  
10: 0.008787530143209422,  
11: 0.009091698338044654,  
12: 0.008505407162453816,  
13: 0.010543631114661873,  
14: 0.011416912465632171,  
15: 0.006866427177335924,  
16: 0.010217528615786823,  
17: 0.007058943147257766,  
18: 0.00865878725453875,  
19: 0.008538083300240204,  
20: 0.010815502968577053,  
21: 0.00842959131917071,  
22: 0.011228866397498058,  
23: 0.009968094879741275,  
24: 0.010770450464362646,  
25: 0.008667254855383425,  
26: 0.010204615665349828,  
27: 0.011372368269527061,  
28: 0.010954880624042367,  
29: 0.009989531452685097,  
30: 0.00962067221335829,  
31: 0.007349942944151046,  
32: 0.008413551765009275,  
33: 0.010223242124173068,  
34: 0.006904986160208653,  
35: 0.0014684288194223183,  
36: 0.009886775554121294,  
37: 0.009349907508568142,  
38: 0.0066854873169042124,  
39: 0.010246940362856517,  
40: 0.009002011751912677,  
41: 0.011335240369516792,  
42: 0.010530113357736223,
```

43: 0.011210363960809685,  
44: 0.01082216518344743,  
45: 0.011208141513571076,  
46: 0.011177811593011785,  
47: 0.010837475774865315,  
48: 0.011351426586467352,  
49: 0.010765387264829987,  
50: 0.011438956678676768,  
51: 0.009361474398626175,  
52: 0.010277764871543938,  
53: 0.011609523940512907,  
54: 0.011246193841519722,  
55: 0.00992529652297925,  
56: 0.011379280475145727,  
57: 0.01049733192901744,  
58: 0.010155052858383277,  
59: 0.007353991424519514,  
60: 0.006804681633064281,  
61: 0.0095365982725504,  
62: 0.0109335789535963,  
63: 0.010817430144035047,  
64: 0.0108678559524874,  
65: 0.008317078780016699,  
66: 0.011381459718355483,  
67: 0.008392962610693304,  
68: 0.009507023410113932,  
69: 0.009172683976940352,  
70: 0.011358015895127065,  
71: 0.010544719984235704,  
72: 0.009393559154450502,  
73: 0.008157143242305242,  
74: 0.006479619610349295,  
75: 0.010619129733755187,  
76: 0.010974899312105888,  
77: 0.00573236844904115,  
78: 0.010308144218856792,  
79: 0.010876717338040125,  
80: 0.00889457890728927,  
81: 0.011384196979315805,  
82: 0.00812596173109734,  
83: 0.011443376271009283,  
84: 0.01053855412592978,  
85: 0.00976458282273997,

```
86: 0.010747089725747434,  
87: 0.008467013021567896,  
88: 0.0106634143792882,  
89: 0.010808975769194117,  
90: 0.008635139461587267,  
91: 0.009748316171145667,  
92: 0.008216121010805459,  
93: 0.010691681343632829,  
94: 0.008792828200159787,  
95: 0.011033746787104561,  
96: 0.010677314513891877,  
97: 0.011775587591147716,  
98: 0.01148635912955692,  
99: 0.011210363960809686,  
100: 0.01082216518344743,  
101: 0.01120814151357108,  
102: 0.011177811593011787}
```

```
In [54]: ranked_sentences = []  
         for i, s in enumerate(sentences):  
             print(i,s)  
             ranked_sentences.append((scores[i],s))  
  
         sorted(ranked_sentences, reverse=True)
```



0 I have bought several of the Vitality canned dog food products and have found them all to be of good quality.

1 The product looks more like a stew than a processed meat and it smells better.

2 My Labrador is finicky and she appreciates this product better than most.

3 Product arrived labeled as Jumbo Salted Peanuts...the peanuts were actually small sized unsalted.

4 Not sure if this was an error or if the vendor intended to represent the product as "Jumbo".

5 This is a confection that has been around a few centuries.

6 It is a light, pillowy citrus gelatin with nuts - in this case Filberts.

7 And it is cut into tiny squares and then liberally coated with powdered sugar.

8 And it is a tiny mouthful of heaven.

9 Not too chewy, and very flavorful.

10 I highly recommend this yummy treat.

11 If you are familiar with the story of C.S.

12 Lewis' "The Lion, The Witch, and The Wardrobe" - this is the treat that seduces Edmund into selling out his Brother and Sisters to the Witch.

13 If you are looking for the secret ingredient in Robitussin I believe I have found it.

14 I got this in addition to the Root Beer Extract I ordered (which was good) and made some cherry soda.

15 The flavor is very medicinal.

16 Great taffy at a great price.

17 There was a wide assortment of yummy taffy.

18 Delivery was very quick.

19 If your a taffy lover, this is a deal.

20 I got a wild hair for taffy and ordered this five pound bag.

21 The taffy was all very enjoyable with many flavors: watermelon, root beer, melon, peppermint, grape, etc.

22 My only complaint is there was a bit too much red/black licorice-flavored pieces (just not my particular favorites).

23 Between me, my kids, and my husband, this lasted only two weeks!

24 I would recommend this brand of taffy -- it was a delightful treat.

25 This saltwater taffy had great flavors and was very soft and chewy.

26 Each candy was individually wrapped well.

27 None of the candies were stuck together, which did happen in the expensive version, Fralinger's.

28 Would highly recommend this candy!

29 I served it at a beach-themed party and everyone loved it!

30 This taffy is so good.

31 It is very soft and chewy.

32 The flavors are amazing.

33 I would definitely recommend you buying it.

34 Very satisfying!

35 !

36 Right now I'm mostly just sprouting this so my cats can eat the grass.

37 They love it.

38 I rotate it around with Wheatgrass and Rye too

39 This is a very healthy dog food.

40 Good for their digestion.  
41 Also good for small puppies.  
42 My dog eats her required amount at every feeding.  
43 I don't know if it's the cactus or the tequila or just the unique combination of ingredients, but the flavor of this hot sauce makes it one of a kind!  
44 We picked up a bottle once on a trip we were on and brought it back home with us and were totally blown away!  
45 When we realized that we simply couldn't find it anywhere in our city we were bummed.<br /><br />Now, because of the magic of the internet, we have a case of the sauce and are ecstatic because of it.<br /><br />If you love hot sauce..I mean really love hot sauce, but don't want a sauce that tastelessly burns your throat, grab a bottle of Tequila Picante Gourmet de Incan.  
46 Just realize that once you taste it, you will never want to use any other sauce.<br /><br />Thank you for the personal, incredible service!  
47 One of my boys needed to lose some weight and the other didn't.  
48 I put this food on the floor for the chubby guy, and the protein-rich, no by-product food up higher where only my skinny boy can jump.  
49 The higher food sits going stale.  
50 They both really go for this food.  
51 And my chubby boy has been losing about an ounce a week.  
52 My cats have been happily eating Felidae Platinum for more than two years.  
53 I just got a new bag and the shape of the food is different.  
54 They tried the new food when I first put it in their bowls and now the bowls sit full and the kitties will not touch the food.  
55 I've noticed similar reviews related to formula changes in the past.  
56 Unfortunately, I now need to find a new food that my cats will eat.  
57 good flavor!  
58 these came securely packed... they were fresh and delicious!  
59 i love these Twizzlers!  
60 The Strawberry Twizzlers are my guilty pleasure - yummy.  
61 Six pounds will be around for a while with my son and I.  
62 My daughter loves twizzlers and this shipment of six pounds really hit the spot.  
63 It's exactly what you would expect...six packages of strawberry twizzlers.  
64 I love eating them and they are good for watching TV and looking at movies!  
65 It is not too sweet.  
66 I like to transfer them to a zip lock baggie so they stay fresh so I can take my time eating them.  
67 I am very satisfied with my Twizzler purchase.  
68 I shared these with others and we have all enjoyed them.  
69 I will definitely be ordering more.  
70 Twizzlers, Strawberry my childhood favorite candy, made in Lancaster Pennsylvania by Y & S Candies, Inc. one of the oldest confectionery Firms in the United States, now a Subsidiary of the Hershey Company, the Company was established in 1845 as Young and Smylie, they also make Apple Licorice Twists, Green Color and Blue Raspberry Licorice Twists, I like them all<br /><br />I keep it in a dry cool place because is not recommended it to put it in the fridge.

71 According to the Guinness Book of Records, the longest Licorice Twist ever made measured 1.200 Feet (370 M) and weighted 100 Pounds (45 Kg) and was made by Y & S Candies, Inc.

72 This Record-Breaking Twist became a Guinness World Record on July 19, 1998.

73 This Product is Kosher!

74 Thank You

75 Candy was delivered very fast and was purchased at a reasonable price.

76 I was home bound and unable to get to a store so this was perfect for me.

77 My husband is a Twizzlers addict.

78 We've bought these many times from Amazon because we're government employees living overseas and can't get them in the country we are assigned to.

79 They've always been fresh and tasty, packed well and arrive in a timely manner.

80 I bought these for my husband who is currently overseas.

81 He loves these, and apparently his staff likes them also.<br />There are generous amounts of Twizzlers in each 16-ounce bag, and this was well worth the price.

82 <a href="http://www.amazon.com/gp/product/B001GVISJM">Twizzlers, Strawberry, 16-Ounce Bags (Pack of 6)</a>

83 I can remember buying this candy as a kid and the quality hasn't dropped in all these years.

84 Still a superb product you won't be disappointed with.

85 I love this candy.

86 After weight watchers I had to cut back but still have a craving for it.

87 I have lived out of the US for over 7 yrs now, and I so miss my Twizzlers!!

88 When I go back to visit or someone visits me, I always stock up.

89 All I can say is YUM!<br />Sell these in Mexico and you will have a faithful buyer, more often than I'm able to buy them right now.

90 Product received is as advertised.<br /><br /><a href="http://www.amazon.com/gp/product/B001GVISJM">Twizzlers, Strawberry, 16-Ounce Bags (Pack of 6)</a>

91 The candy is just red , No flavor .

92 Just plan and chewy .

93 I would never buy them again

94 I was so glad Amazon carried these batteries.

95 I have a hard time finding them elsewhere because they are such a unique size.

96 I need them for my garage door opener.<br />Great deal for the price.

97 I got this for my Mum who is not diabetic but needs to watch her sugar intake, and my father who simply chooses to limit unnecessary sugar intake - she's the one with the sweet tooth - they both LOVED these toffees, you would never guess that they're sugar-free and it's so great that you can eat them pretty much guilt free!

98 i was so impressed that i've ordered some for myself (w dark chocolate) to take to the office so i'll eat them instead of snacking on sugary sweets.<br />These are just EXCELLENT!

99 I don't know if it's the cactus or the tequila or just the unique combination of ingredients, but the flavor of this hot sauce makes it one of a kind!

100 We picked up a bottle once on a trip we were on and brought it back home with us and were totally blown away!

101 When we realized that we simply couldn't find it anywhere in our city we were bummed.<br /><br />Now, because of the magic of the internet, we have a case of the sauce and are ecstatic because of it.<br /><br />If you love hot sauce..I mean really love hot sauce, but don't want a sauce that tastelessly burns your throat,

grab a bottle of Tequila Picante Gourmet de Inclan.

102 Just realize that once you taste it, you will never want to use any other sauce.<br /><br />Thank you for the personal, incredible service!

```

Out[54]: [(0.011775587591147716,
    "I got this for my Mum who is not diabetic but needs to watch her sugar intake, and my father who simply chooses to limit unnecessary sugar intake - she's the one with the sweet tooth - they both LOVED these toffees, you would never guess that they're sugar-free and it's so great that you can eat them pretty much guilt free!",
    (0.011609523940512907,
    'I just got a new bag and the shape of the food is different.'),
    (0.01148635912955692,
    "i was so impressed that i've ordered some for myself (w dark chocolate) to take to the office so i'll eat them instead of snacking on sugary sweets.<br />These are just EXCELLENT!"),
    (0.011443376271009283,
    "I can remember buying this candy as a kid and the quality hasn't dropped in all these years."),
    (0.011438956678676768, 'They both really go for this food.'),
    (0.011416912465632171,
    'I got this in addition to the Root Beer Extract I ordered (which was good) and made some cherry soda.'),
    (0.011384196979315805,
    'He loves these, and apparently his staff likes them also.<br />There are generous amounts of Twizzlers in each 16-ounce bag, and this was well worth the price.'),
    (0.011381459718355483,
    'I like to transfer them to a zip lock baggie so they stay fresh so I can take my time eating them.'),
    (0.011379280475145727,
    'Unfortunately, I now need to find a new food that my cats will eat.'),
    (0.011372368269527061,
    "None of the candies were stuck together, which did happen in the expensive version, Fralinger's."),
    (0.011363842644330402,
    'I have bought several of the Vitality canned dog food products and have found them all to be of good quality.'),
    (0.011358015895127065,
    'Twizzlers, Strawberry my childhood favorite candy, made in Lancaster Pennsylvania by Y & S Candies, Inc. one of the oldest confectionery Firms in the United States, now a Subsidiary of the Hershey Company, the Company was established in 1845 as Young and Smylie, they also make Apple Licorice Twists, Green Color and Blue Raspberry Licorice Twists, I like them all<br /><br />I keep it in a dry cool place because is not recommended it to put it in the fridge.'),
    (0.011351426586467352,
    'I put this food on the floor for the chubby guy, and the protein-rich, no by-product food up higher where only my skinny boy can jump.'),
    (0.011335240369516792, 'Also good for small puppies.'),
    (0.011246193841519722,
    'They tried the new food when I first put it in their bowls and now the bowls sit full and the kitties will not touch the food.'),
    (0.011228866397498058,
    'My only complaint is there was a bit too much red/black licorice-flavored pieces (just not my particular favorites).'),

```

(0.011210363960809686,  
"I don't know if it's the cactus or the tequila or just the unique combination of ingredients, but the flavor of this hot sauce makes it one of a kind!"),  
(0.011210363960809685,  
"I don't know if it's the cactus or the tequila or just the unique combination of ingredients, but the flavor of this hot sauce makes it one of a kind!"),  
(0.01120814151357108,  
"When we realized that we simply couldn't find it anywhere in our city we were bummed.<br /><br />Now, because of the magic of the internet, we have a case of the sauce and are ecstatic because of it.<br /><br />If you love hot sauce..I mean really love hot sauce, but don't want a sauce that tastelessly burns your throat, grab a bottle of Tequila Picante Gourmet de Inclan."),  
(0.011208141513571076,  
"When we realized that we simply couldn't find it anywhere in our city we were bummed.<br /><br />Now, because of the magic of the internet, we have a case of the sauce and are ecstatic because of it.<br /><br />If you love hot sauce..I mean really love hot sauce, but don't want a sauce that tastelessly burns your throat, grab a bottle of Tequila Picante Gourmet de Inclan."),  
(0.011177811593011787,  
'Just realize that once you taste it, you will never want to use any other sauce.<br /><br />Thank you for the personal, incredible service!'),  
(0.011177811593011785,  
'Just realize that once you taste it, you will never want to use any other sauce.<br /><br />Thank you for the personal, incredible service!'),  
(0.011140570518279798,  
'The product looks more like a stew than a processed meat and it smells better.'),  
(0.011033746787104561,  
'I have a hard time finding them elsewhere because they are such a unique size.'),  
(0.010974899312105888,  
'I was home bound and unable to get to a store so this was perfect for me.'),  
(0.010954880624042367, 'Would highly recommend this candy!'),  
(0.0109335789535963,  
'My daughter loves twizzlers and this shipment of six pounds really hit the spot.'),  
(0.010876717338040125,  
"They've always been fresh and tasty, packed well and arrive in a timely manner."),  
(0.0108678559524874,  
'I love eating them and they are good for watching TV and looking at movies!'),  
(0.010837475774865315,  
"One of my boys needed to lose some weight and the other didn't."),  
(0.01082216518344743,  
'We picked up a bottle once on a trip we were on and brought it back home with us and were totally blown away!'),  
(0.01082216518344743,  
'We picked up a bottle once on a trip we were on and brought it back home with us and were totally blown away!'),

(0.010817430144035047,  
"It's exactly what you would expect...six packages of strawberry twizzlers."),  
(0.010815502968577053,  
'I got a wild hair for taffy and ordered this five pound bag.'),  
(0.010808975769194117,  
"All I can say is YUM!<br />Sell these in Mexico and you will have a faithful buyer, more often than I'm able to buy them right now."),  
(0.010770450464362646,  
'I would recommend this brand of taffy -- it was a delightful treat.'),  
(0.010765387264829987, 'The higher food sits going stale.'),  
(0.010747089725747434,  
'After weight watchers I had to cut back but still have a craving for it.'),  
(0.010691681343632829, 'I would never buy them again'),  
(0.010677314513891877,  
'I need them for my garage door opener.<br />Great deal for the price.'),  
(0.0106634143792882,  
'When I go back to visit or someone visits me, I always stock up.'),  
(0.010619129733755187,  
'Candy was delivered very fast and was purchased at a reasonable price.'),  
(0.010544719984235704,  
'According to the Guinness Book of Records, the longest Licorice Twist ever made measured 1.200 Feet (370 M) and weighted 100 Pounds (45 Kg) and was made by Y & S Candies, Inc.'),  
(0.010543631114661873,  
'If you are looking for the secret ingredient in Robitussin I believe I have found it.'),  
(0.01053855412592978,  
"Still a superb product you won't be disappointed with."),  
(0.010530113357736223, 'My dog eats her required amount at every feeding.'),  
(0.01049733192901744, 'good flavor!'),  
(0.01033199595784022,  
'Not sure if this was an error or if the vendor intended to represent the product as "Jumbo".'),  
(0.010308144218856792,  
"We've bought these many times from Amazon because we're government employees living overseas and can't get them in the country we are assigned to."),  
(0.010277764871543938,  
'My cats have been happily eating Felidae Platinum for more than two years.'),  
(0.010246940362856517, 'This is a very healthy dog food.'),  
(0.010223242124173068, 'I would definitely recommend you buying it.'),  
(0.010217528615786823, 'Great taffy at a great price.'),  
(0.010204615665349828, 'Each candy was individually wrapped well.'),  
(0.010155052858383277,  
'these came securely packed... they were fresh and delicious!'),  
(0.009989531452685097,  
'I served it at a beach-themed party and everyone loved it!'),

```
(0.009968094879741275,
'Between me, my kids, and my husband, this lasted only two weeks!'),
(0.00992529652297925,
"I've noticed similar reviews related to formula changes in the past."),
(0.009886775554121294,
"Right now I'm mostly just sprouting this so my cats can eat the grass."),
(0.00976458282273997, 'I love this candy.'),
(0.009748316171145667, 'The candy is just red , No flavor .'),
(0.009659130193868583,
'Product arrived labeled as Jumbo Salted Peanuts...the peanuts were actually small sized unsalted.'),
(0.00962067221335829, 'This taffy is so good.'),
(0.0095365982725504,
'Six pounds will be around for a while with my son and I.'),
(0.009507023410113932,
'I shared these with others and we have all enjoyed them.'),
(0.009393559154450502,
'This Record-Breaking Twist became a Guinness World Record on July 19, 1998.'),
(0.009361474398626175,
'And my chubby boy has been losing about an ounce a week.'),
(0.009349907508568142, 'They love it.'),
(0.009217590489292707,
'My Labrador is finicky and she appreciates this product better than most.'),
(0.009172683976940352, 'I will definitely be ordering more.'),
(0.009091698338044654, 'If you are familiar with the story of C.S.'),
(0.009002011751912677, 'Good for their digestion.'),
(0.00889457890728927,
'I bought these for my husband who is currently overseas.'),
(0.008792828200159787, 'I was so glad Amazon carried these batteries.'),
(0.008787530143209422, 'I highly recommend this yummy treat.'),
(0.008667254855383425,
'This saltwater taffy had great flavors and was very soft and chewy.'),
(0.00865878725453875, 'Delivery was very quick.'),
(0.008635139461587267,
'Product received is as advertised.<br /><br /><a href="http://www.amazon.com/gp/product/B001GVISJM">Twizzl
ers, Strawberry, 16-Ounce Bags (Pack of 6)</a>'),
(0.008538083300240204, 'If your a taffy lover, this is a deal.'),
(0.00853471941472121,
'This is a confection that has been around a few centuries.'),
(0.008505407162453816,
'Lewis\' "The Lion, The Witch, and The Wardrobe" - this is the treat that seduces Edmund into selling out h
is Brother and Sisters to the Witch.'),
(0.008467013021567896,
'I have lived out of the US for over 7 yrs now, and I so miss my Twizzlers!!'),
```



```
(0.00842959131917071,
 'The taffy was all very enjoyable with many flavors: watermelon, root beer, melon, peppermint, grape, et
c. '),
(0.008413551765009275, 'The flavors are amazing.'),
(0.008392962610693304, 'I am very satisfied with my Twizzler purchase.'),
(0.008351750313671847,
 'It is a light, pillowy citrus gelatin with nuts - in this case Filberts.'),
(0.008346515223468891,
 'And it is cut into tiny squares and then liberally coated with powdered sugar.'),
(0.008317078780016699, 'It is not too sweet.'),
(0.008216121010805459, 'Just plan and chewy .'),
(0.008157143242305242, 'This Product is Kosher!'),
(0.00812596173109734,
 '<a href="http://www.amazon.com/gp/product/B001GVISJM">Twizzlers, Strawberry, 16-Ounce Bags (Pack of 6)</a
>'),
(0.008123595327268229, 'And it is a tiny mouthful of heaven.'),
(0.007353991424519514, 'i love these Twizzlers!'),
(0.007349942944151046, 'It is very soft and chewy.'),
(0.007058943147257766, 'There was a wide assortment of yummy taffy.'),
(0.006904986160208653, 'Very satisfying!'),
(0.006866427177335924, 'The flavor is very medicinal.'),
(0.006804681633064281,
 'The Strawberry Twizzlers are my guilty pleasure - yummy.'),
(0.0066854873169042124, 'I rotate it around with Wheatgrass and Rye too'),
(0.006479619610349295, 'Thank You'),
(0.00573236844904115, 'My husband is a Twizzlers addict.'),
(0.0047202546527219495, 'Not too chewy, and very flavorful.'),
(0.0014684288194223183, '!')]
```

```
In [58]: for i in range(1):
          print('-----')
          print("Entire Text:")
          print(df['Text'][i])
          print('\n')
          print("Original Summary:")
          print(df['Summary'][i])
          print('\n')
          print("Predicted Summary:")
          print(ranked_sentences[i][1]+ranked_sentences[i+1][1])
          print('\n')
```

-----

Entire Text:

I have bought several of the Vitality canned dog food products and have found them all to be of good quality. The product looks more like a stew than a processed meat and it smells better. My Labrador is finicky and she appreciates this product better than most.

Original Summary:

Good Quality Dog Food

Predicted Summary:

I have bought several of the Vitality canned dog food products and have found them all to be of good quality. The product looks more like a stew than a processed meat and it smells better.

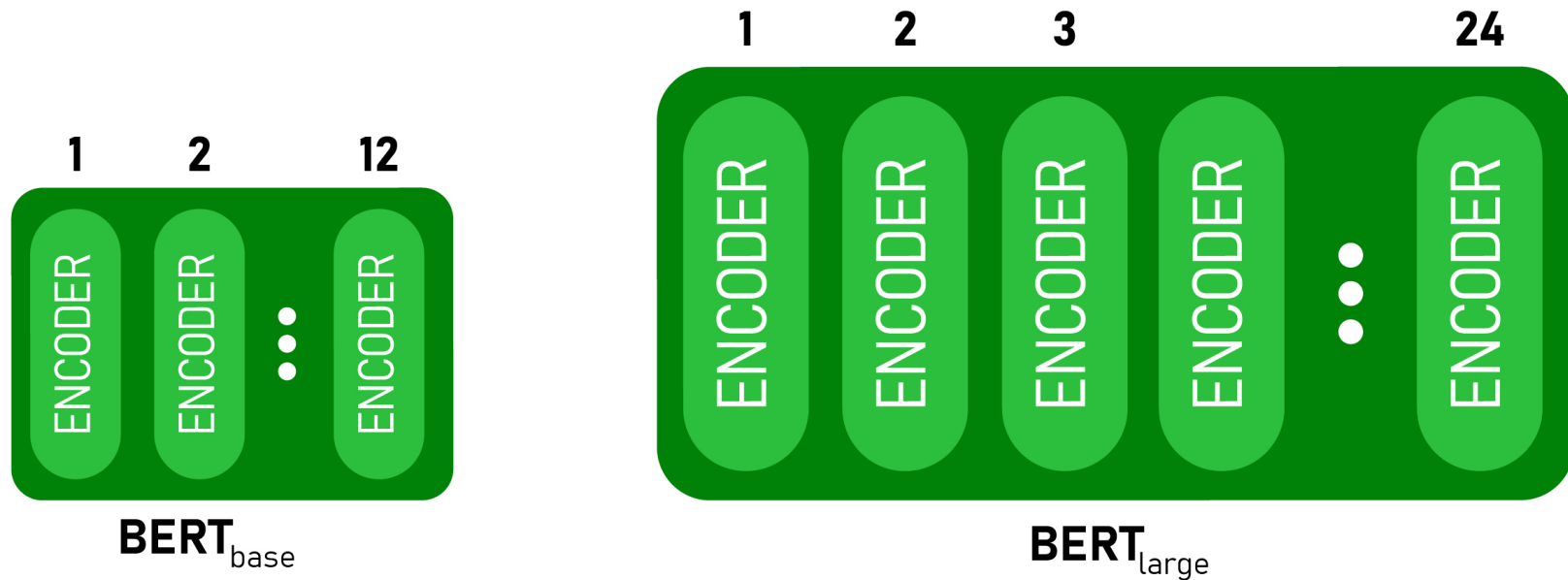
## Abstractive Approach

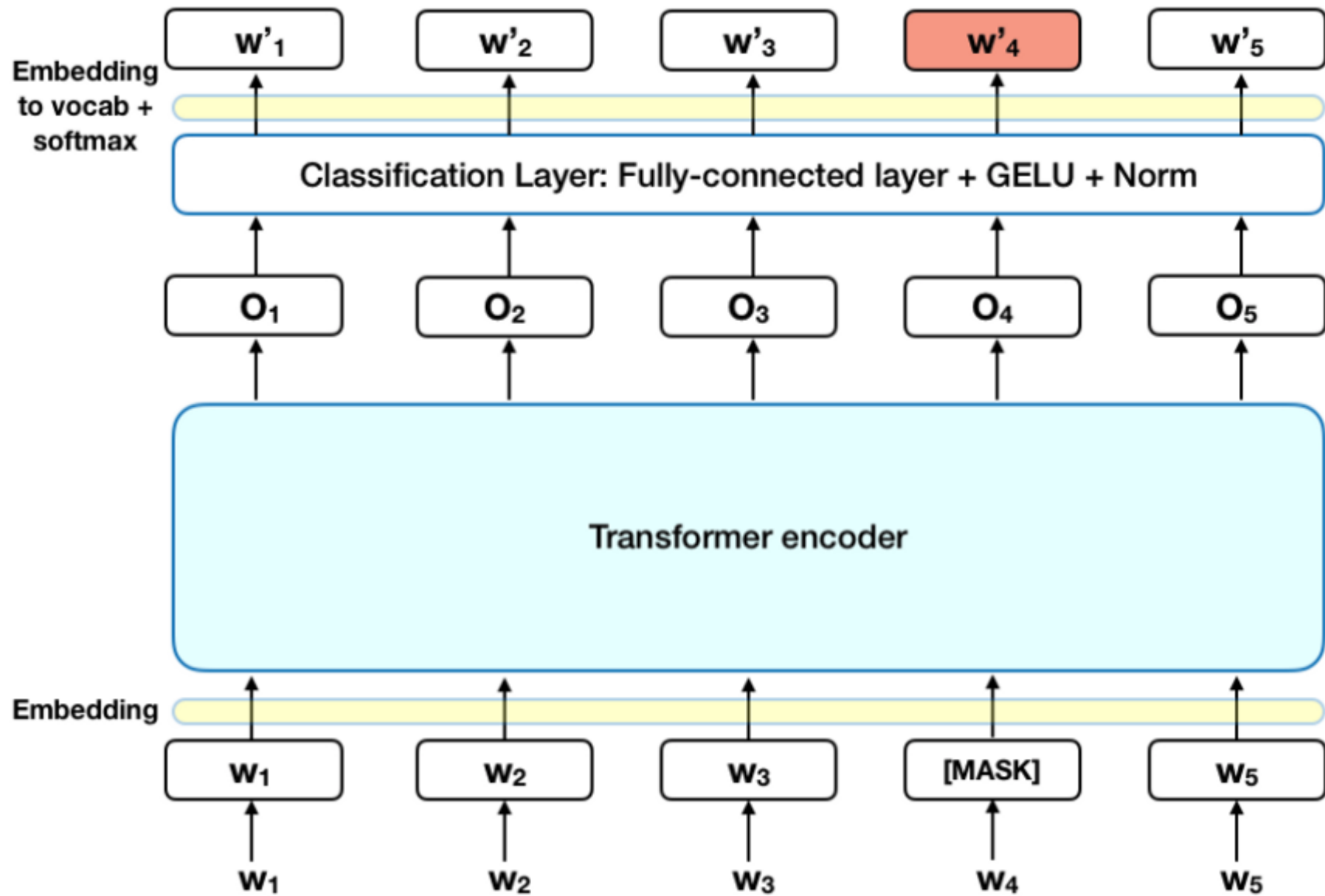
Text Summarization using using BERT, GPT2, XLNET

## Model 2 - BERT (Bidirectional Encoder Representations from Transformers) Model

BERT makes use of Transformer, an attention mechanism that learns contextual relations between words (or sub-words) in a text. In its vanilla form, Transformer includes two separate mechanisms — an encoder that reads the text input and a decoder that produces a prediction for the task. Since BERT's goal is to generate a language model, only the encoder mechanism is necessary.

Word Embeddings Used = ELMo





```
In [ ]: from summarizer import Summarizer, TransformerSummarizer

for i in range(5):
    print('-----')
    print("Entire Text:")
    print(df['Text'][i])
    print('\n')
    print("Original Summary:")
    print(df['Summary'][i])
    print('\n')
    bert_model = Summarizer()
    bert_summary = ''.join(bert_model(df['Text'][i], min_length=2))
    print("Predicted Summary:")
    print(bert_summary)
    print('\n')
```

-----  
Entire Text:

I have bought several of the Vitality canned dog food products and have found them all to be of good quality. The product looks more like a stew than a processed meat and it smells better. My Labrador is finicky and she appreciates this product better than most.

Original Summary:

Good Quality Dog Food

Predicted Summary:

I have bought several of the Vitality canned dog food products and have found them all to be of good quality. The product looks more like a stew than a processed meat and it smells better.

-----  
Entire Text:

Product arrived labeled as Jumbo Salted Peanuts...the peanuts were actually small sized unsalted. Not sure if this was an error or if the vendor intended to represent the product as "Jumbo".

Original Summary:

Not as Advertised

Predicted Summary:

Product arrived labeled as Jumbo Salted Peanuts...the peanuts were actually small sized unsalted.

-----  
Entire Text:

This is a confection that has been around a few centuries. It is a light, pillowy citrus gelatin with nuts - in this case Filberts. And it is cut into tiny squares and then liberally coated with powdered sugar. And it is a tiny mouthful of heaven. Not too chewy, and very flavorful. I highly recommend this yummy treat. If you are familiar with the story of C.S. Lewis' "The Lion, The Witch, and The Wardrobe" - this is the treat that seduces Edmund into selling out his Brother and Sisters to the Witch.

Original Summary:

"Delight" says it all

Predicted Summary:

This is a confection that has been around a few centuries. Not too chewy, and very flavorful.

-----

Entire Text:

If you are looking for the secret ingredient in Robitussin I believe I have found it. I got this in addition to the Root Beer Extract I ordered (which was good) and made some cherry soda. The flavor is very medicinal.

Original Summary:

Cough Medicine

Predicted Summary:

If you are looking for the secret ingredient in Robitussin I believe I have found it. The flavor is very medicinal.

-----

Entire Text:

Great taffy at a great price. There was a wide assortment of yummy taffy. Delivery was very quick. If your a taffy lover, this is a deal.

Original Summary:

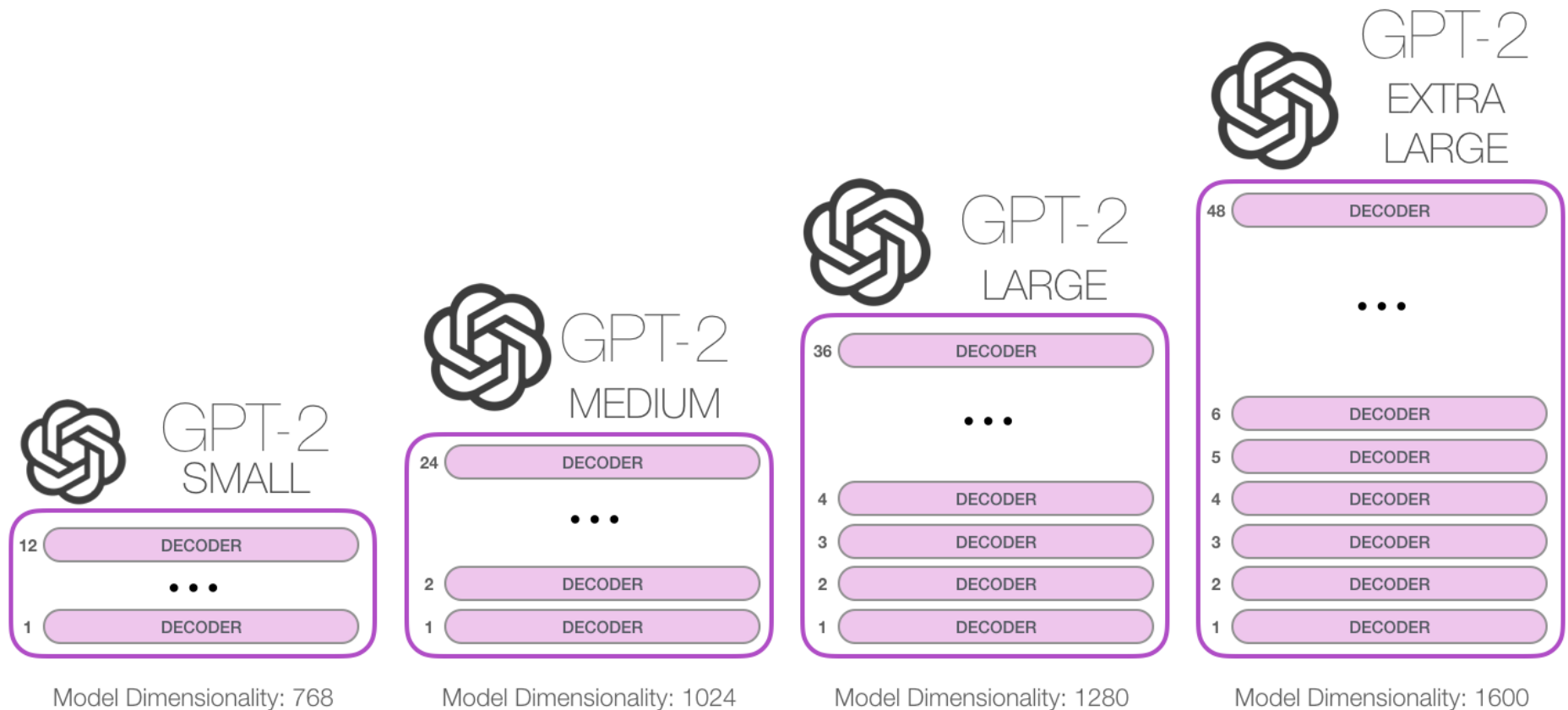
Great taffy

Predicted Summary:

Great taffy at a great price. If your a taffy lover, this is a deal.

### Model 3 - GPT-2 (Generative Pre-trained Transformer) Model

GPT is a transformer model, which uses "attention" in place of previous recurrence- and convolution-based architectures. Attention mechanisms allow the model to selectively focus on segments of input text it predicts to be the most relevant.





```
In [ ]: for i in range(5):
        print('-----')
        print("Entire Text:")
        print(df['Text'][i])
        print('\n')
        print("Original Summary:")
        print(df['Summary'][i])
        print('\n')
        GPT2_model = TransformerSummarizer(transformer_type="GPT2",transformer_model_key="gpt2-medium")
        full = ''.join(GPT2_model(df['Text'][i], min_length=2))
        print("Predicted Summary:")
        print(full)
        print('\n')
```

-----  
Entire Text:

I have bought several of the Vitality canned dog food products and have found them all to be of good quality. The product looks more like a stew than a processed meat and it smells better. My Labrador is finicky and she appreciates this product better than most.

Original Summary:

Good Quality Dog Food

Predicted Summary:

I have bought several of the Vitality canned dog food products and have found them all to be of good quality.

-----  
Entire Text:

Product arrived labeled as Jumbo Salted Peanuts...the peanuts were actually small sized unsalted. Not sure if this was an error or if the vendor intended to represent the product as "Jumbo".

Original Summary:

Not as Advertised

Predicted Summary:

Product arrived labeled as Jumbo Salted Peanuts...the peanuts were actually small sized unsalted. Not sure if this was an error or if the vendor intended to represent the product as "Jumbo".

-----  
Entire Text:

This is a confection that has been around a few centuries. It is a light, pillowy citrus gelatin with nuts - in this case Filberts. And it is cut into tiny squares and then liberally coated with powdered sugar. And it is a tiny mouthful of heaven. Not too chewy, and very flavorful. I highly recommend this yummy treat. If you are familiar with the story of C.S. Lewis' "The Lion, The Witch, and The Wardrobe" - this is the treat that seduces Edmund into selling out his Brother and Sisters to the Witch.

Original Summary:

"Delight" says it all

Predicted Summary:

This is a confection that has been around a few centuries.

-----

Entire Text:

If you are looking for the secret ingredient in Robitussin I believe I have found it. I got this in addition to the Root Beer Extract I ordered (which was good) and made some cherry soda. The flavor is very medicinal.

Original Summary:

Cough Medicine

Predicted Summary:

If you are looking for the secret ingredient in Robitussin I believe I have found it.

-----

Entire Text:

Great taffy at a great price. There was a wide assortment of yummy taffy. Delivery was very quick. If your a taffy lover, this is a deal.

Original Summary:

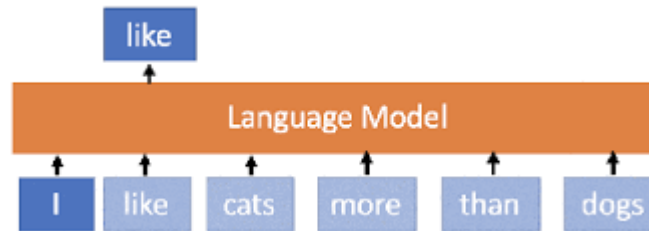
Great taffy

Predicted Summary:

Great taffy at a great price.

### Model 4 - XLNet Model

XLNet is a BERT-like model instead of a totally different one. It is a kind of model that using the context word to predict the next word. But here the context word is constrained to two directions, either forward or backward.



```
In [ ]: for i in range(5):
        print('-----')
        print("Entire Text:")
        print(df['Text'][i])
        print('\n')
        print("Original Summary:")
        print(df['Summary'][i])
        print('\n')
        model = TransformerSummarizer(transformer_type="XLNet",transformer_model_key="xlnet-base-cased")
        full = ''.join(model(df['Text'][i], min_length=2))
        print("Predicted Summary:")
        print(full)
        print('\n')
```

-----  
Entire Text:

I have bought several of the Vitality canned dog food products and have found them all to be of good quality. The product looks more like a stew than a processed meat and it smells better. My Labrador is finicky and she appreciates this product better than most.

Original Summary:

Good Quality Dog Food

Predicted Summary:

I have bought several of the Vitality canned dog food products and have found them all to be of good quality. My Labrador is finicky and she appreciates this product better than most.

-----  
Entire Text:

Product arrived labeled as Jumbo Salted Peanuts...the peanuts were actually small sized unsalted. Not sure if this was an error or if the vendor intended to represent the product as "Jumbo".

Original Summary:

Not as Advertised

Predicted Summary:

Product arrived labeled as Jumbo Salted Peanuts...the peanuts were actually small sized unsalted.

-----  
Entire Text:

This is a confection that has been around a few centuries. It is a light, pillowy citrus gelatin with nuts - in this case Filberts. And it is cut into tiny squares and then liberally coated with powdered sugar. And it is a tiny mouthful of heaven. Not too chewy, and very flavorful. I highly recommend this yummy treat. If you are familiar with the story of C.S. Lewis' "The Lion, The Witch, and The Wardrobe" - this is the treat that seduces Edmund into selling out his Brother and Sisters to the Witch.

Original Summary:

"Delight" says it all

Predicted Summary:

This is a confection that has been around a few centuries. It is a light, pillowy citrus gelatin with nuts - in this case Filberts.

-----

Entire Text:

If you are looking for the secret ingredient in Robitussin I believe I have found it. I got this in addition to the Root Beer Extract I ordered (which was good) and made some cherry soda. The flavor is very medicinal.

Original Summary:

Cough Medicine

Predicted Summary:

If you are looking for the secret ingredient in Robitussin I believe I have found it.

-----

Entire Text:

Great taffy at a great price. There was a wide assortment of yummy taffy. Delivery was very quick. If your a taffy lover, this is a deal.

Original Summary:

Great taffy

Predicted Summary:

Great taffy at a great price.

## Conclusion

1. The Extractive Model works well for only limited number of reviews. With an increase the number of rows in the dataset, the complexity to find the similarity matrix increases exponentially.
2. Modern Abstractive Models with LSTM based attention transformers are best suited for such kind of applications. Since, they are good at predicting the next word or good at extracting specific target words (to be retained in the end) which helps to summarize based on predicted model values.
3. Among all the models, the XLNet gave the most accurate and consise summaries for almost all the reviews.

```
In [ ]: [!]jupyter nbconvert --to html "ML_Experiment_6.ipynb"
```

```
[NbConvertApp] Converting notebook ML_Experiment_6.ipynb to html
```