Stance Detection for the Fake News Challenge

Identifying Textual Relationships with Deep Neural Nets

Check the problem context here.

Download files required for the project from here.

Step1: Load the given dataset

1. Mount the google drive
2. Import Glove embeddings
3. Import the test and train datasets

Mount the google drive to access required project files

Run the below commands

```
from google.colab import drive
```

```
drive.mount('/content/drive/')
```

```
Go to this URL in a browser: https://accounts.google.com/o/oauth2/auth?client_id=947318989803-6

Enter your authorization code:
··········
Mounted at /content/drive/
```

Path for Project files on google drive

Note: You need to change this path according where you have kept the files in google drive.

```
project_path = "/content/drive/My Drive/Colab Notebooks/Sequence NLP/stance detection/"
```

Loading the Glove Embeddings

```
from zipfile import ZipFile
with ZipFile(project_path+'glove.6B.zip', 'r') as z:
  z.extractall()
```

Load the dataset [5 Marks]

1. Using read_csv() in pandas load the given train datasets files `train_bodies.csv` and `train_stances.cs`
2. Using merge command in pandas merge the two datasets based on the Body ID.

Note: Save the final merged dataset in a dataframe with name `dataset`.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline

train_bodies = pd.read_csv(project_path+'train_bodies.csv')
train_instances = pd.read_csv(project_path+'train_stances.csv')
```

```
train_bodies.head()
```

	Body ID	articleBody
0	0	A small meteorite crashed into a wooded area i...
1	4	Last week we hinted at what was to come as Ebo...
2	5	(NEWSER) – Wonder how long a Quarter Pounder w...
3	6	Posting photos of a gun-toting child online, I...
4	7	At least 25 suspected Boko Haram insurgents we...

```
train_instances.head()
```

	Headline	Body ID	Stance
0	Police find mass graves with at least '15 bodi...	712	unrelated
1	Hundreds of Palestinians flee floods in Gaza a...	158	agree
2	Christian Bale passes on role of Steve Jobs, a...	137	unrelated
3	HBO and Apple in Talks for $15/Month Apple TV ...	1034	unrelated
4	Spider burrowed through tourist's stomach and ...	1923	disagree

```
# Merge both the dataframes
dataset = pd.merge(train_bodies, train_instances,on ='Body ID')
```

Check1:

You should see the below output if you run `dataset.head()` command as given below

```
dataset.head()
```

	Body ID	articleBody	Headlin
0	0	A small meteorite crashed into a wooded area i...	Soldier shot, Parliament locked down after gun.
1	0	A small meteorite crashed into a wooded area i...	Tourist dubbed ‘Spider Man’ after spider burro.
2	0	A small meteorite crashed into a wooded area i...	Luke Somers 'killed in failed rescue attempt i.
3	0	A small meteorite crashed into a wooded area i...	BREAKING: Soldier shot at War Memorial in Ottaw
4	0	A small meteorite crashed into a wooded area i...	Giant 8ft 9in catfish weighing 19 stone caught.

Step2: Data Pre-processing and setting some hyper parameters needed for

Run the code given below to set the required parameters.

1. `MAX_SENTS` = Maximum no.of sentences to consider in an article.
2. `MAX_SENT_LENGTH` = Maximum no.of words to consider in a sentence.
3. `MAX_NB_WORDS` = Maximum no.of words in the total vocabualry.
4. `MAX_SENTS_HEADING` = Maximum no.of sentences to consider in a heading of an article.

```
MAX_NB_WORDS = 20000
MAX_SENTS = 20
MAX_SENTS_HEADING = 1
MAX_SENT_LENGTH = 20
VALIDATION_SPLIT = 0.2
```

Download the `Punkt` from nltk using the commands given below. This is for sentence t

For more info on how to use it, read this.

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

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt.zip.
True
```

Tokenizing the text and loading the pre-trained Glove word embedd [5 marks]

Keras provides Tokenizer API for preparing text. Read it before going any further.

Import the Tokenizer from keras preprocessing text

```
from keras.preprocessing.text import Tokenizer
```

```
Using TensorFlow backend.
```

Initialize the Tokenizer class with maximum vocabulary count as `MAX_NB_WORDS` initialized at the

```
tokenizer = Tokenizer(num_words=MAX_NB_WORDS )
```

Now, using fit_on_texts() from Tokenizer class, lets encode the data

Note: We need to fit articleBody and Headline also to cover all the words.

```
codetext = dataset['articleBody'] + dataset['Headline']
tokenizer.fit_on_texts(codetext)
```

```
tokenizer.word_counts
```

```
OrderedDict([('a', 478453),
             ('small', 5225),
             ('meteorite', 5773),
             ('crashed', 207),
             ('into', 25604),
             ('wooded', 554),
             ('area', 7132),
             ('in', 408065),
             ("nicaragua's", 526),
             ('capital', 6867),
             ('of', 441773),
             ('managua', 1823),
             ('overnight', 749),
             ('the', 1129038),
             ('government', 23631),
             ('said', 134456),
             ('sunday', 5416),
             ('residents', 3428),
             ('reported', 21926),
             ('hearing', 2514),
             ('mysterious', 853),
             ('boom', 1459),
             ('that', 234408),
             ('left', 11643),
             ('16', 2172),
             ('foot', 2441),
             ('deep', 1218),
             ('crater', 3683),
             ('near', 12938),
             ("city's", 726),
             ('airport', 5586),
             ('associated', 4582),
             ('press', 9235),
             ('reports', 24310),
             ('spokeswoman', 2460),
             ('rosario', 491),
             ('murillo', 1012),
             ('committee', 2417),
             ('formed', 650),
             ('by', 104742),
             ('to', 499718),
             ('study', 1432),
             ('event', 4821),
             ('determined', 1234),
             ('it', 125900),
             ('was', 180053),
             ('relatively', 703),
             ('appears', 7092),
             ('have', 110414),
             ('come', 8558),
             ('off', 15561),
             ('an', 91180),
             ('asteroid', 2700),
             ('passing', 1254),
             ('close', 4953),
             ('earth', 3332),
             ('house', 11703),
             ('sized', 574),
             ('2014', 10895),
             ('rc', 903),
             ('which', 40671),
             ('measured', 228),
             ('60', 1487),
             ('feet', 2115),
             ('diameter', 71),
             ('skimmed', 36),
             ('this', 69775),
             ('weekend', 3506),
             ('abc', 508),
             ('news', 29545),
             ('nicaragua', 1614),
             ('will', 44647),
             ('ask', 2405),
             ('international', 6863),
             ('experts', 3327),
             ('help', 6673),
             ('local', 8747),
             ('scientists', 1441),
             ('understanding', 432),
             ('what', 28417),
             ('happened', 5157),
             ('had', 69250),
             ('radius', 198),
             ('39', 1067),
             ('and', 383581),
             ('depth', 335),
             ('humberto', 638),
             ('saballos', 346),
             ('volcanologist', 332),
             ('with', 129916),
             ('nicaraguan', 2408),
             ('institute', 2007),
             ('territorial', 714),
             ('studies', 1537),
             ('who', 70648),
             ('on', 179598),
             ('he', 131558),
             ('is', 188509),
             ('still', 13329),
             ('not', 82447),
             ('clear', 4723),
             ('if', 30354),
             ('disintegrated', 555),
             ('or', 41957),
             ('buried', 1384),
             ('garcia', 297),
             ('astronomy', 287),
             ('center', 4178),
             ('at', 97873),
             ('national', 7634),
             ('autonomous', 622),
             ('university', 3253),
             ('could', 28627),
             ('be', 93817),
             ('related', 3293),
             ('forecast', 467),
             ('pass', 1102),
             ('planet', 519),
             ('saturday', 4821),
             ('night', 8348),
             ('we', 48610),
             ('more', 36062),
             ('because', 17094),
             ('ice', 373),
             ('rock', 1268),
             ('wilfried', 394),
             ('strauch', 643),
             ('adviser', 511),
             ('very', 11340),
             ('strange', 1843),
             ('no', 34454),
             ('one', 42775),
             ('streak', 281),
             ('light', 1918),
             ('anyone', 5022),
             ('has', 99994),
             ('photo', 9041),
             ('something', 8614),
             ('loud', 1425),
             ('but', 70970),
             ('they', 61196),
             ("didn't", 3678),
             ('see', 12601),
             ('anything', 2902),
             ('sky', 1183),
             ('i', 48949),
             ('sitting', 1198),
             ('my', 18011),
             ('porch', 290),
             ('saw', 4391),
             ('nothing', 3904),
             ('then', 20438),
             ('all', 34207),
             ('sudden', 970),
             ('heard', 9335),
             ('large', 4992),
             ('blast', 1325),
             ('thought', 7011),
             ('bomb', 1183),
             ('felt', 2933),
             ('expansive', 432),
             ('wave', 892),
             ('jorge', 455),
             ('santamaria', 373),
             ('told', 35463),
             ('site', 8965),
             ("managua's", 514),
             ('air', 13072),
             ('force', 5416),
             ('base', 1305),
             ('only', 16209),
             ('journalists', 2837),
             ('from', 97677),
             ('state', 38464),
             ('media', 21144),
             ('were', 48691),
             ('allowed', 1405),
             ('visit', 2180),
             ('soldier', 3534),
             ('shot', 10754),
             ('parliament', 6986),
             ('locked', 211),
             ('down', 11448),
             ('after', 53153),
             ('gunfire', 1968),
             ('erupts', 93),
             ('war', 9660),
             ('memorial', 3741),
             ('tourist', 146),
             ('dubbed', 594),
             ('‘spider', 30),
             ("'spider", 50),
             ('man’', 49),
             ('spider', 6283),
             ('burrows', 248),
             ('under', 12945),
             ('skin', 4315),
             ('for', 147708),
             ('days', 10649),
             ('luke', 1173),
             ('somers', 2614),
             ("'killed", 124),
             ('failed', 2110),
             ('rescue', 2027),
             ('attempt', 1892),
             ("yemen'", 70),
             ('breaking', 1558),
             ('ottawa', 5033),
             ('giant', 726),
             ('8ft', 76),
             ('9in', 36),
             ('catfish', 2432),
             ('weighing', 318),
             ('19', 2243),
             ('stone', 1259),
             ('caught', 5582),
             ('italy', 765),
             ('biggest', 1392),
             ('ever', 4920),
             ('reeled', 35),
             ('rod', 236),
             ('line', 2940),
             ('enormous', 600),
             ('20', 5162),
             ('fishing', 485),
             ('40', 4481),
             ('minute', 1740),
             ('boat', 652),
             ('battle', 1670),
             ('italian', 1030),
             ('catches', 309),
             ('huge', 2223),
             ('wels', 436),
             ('record', 2400),
             ('coming', 5890),
             ('store', 6214),
             ('you', 31209),
             ('pumpkin', 3663),
             ('spice', 4121),
             ('condom', 1964),
             ('gunman', 2851),
             ('killed', 17745),
             ('shooting', 11369),
             ('hill', 3162),
             ('hunt', 595),
             ('other', 21273),
             ('shooters', 124),
             ('canada', 1839),
             ('surreal', 48),
             ('photos', 3936),
             ('fisherman’s', 24),
             ('jaw', 37),
             ('dropping', 307),
             ('catch', 843),
             ('likely', 6320),
             ('people', 33086),
             ('people', 33086),
             ('wondering', 395),
             ('it’s', 11246),
             ('real', 7749),
             ('fisherman', 274),
             ('lands', 138),
             ('world', 13113),
             ('hooked', 440),
             ('source', 10357),
             ('tom', 836),
             ('brokaw', 413),
             ('wants', 2219),
             ('brian', 1135),
             ('williams', 1968),
             ('fired', 6723),
             ('been', 75042),
             ('canada’s', 278),
             ('just', 24249),
             ('steps', 946),
             ('away', 6508),
             ('nation’s', 353),
             ('280', 169),
             ('pound', 621),
             ('makes', 3007),
             ('set', 7532),
             ('rumor', 3468),
             ('debunked', 618),
             ('robocop', 124),
             ('style', 1050),
             ('robots', 1495),
             ('are', 76071),
             ('patrolling', 285),
             ("microsoft's", 504),
             ('campus', 1370),
             ('po', 721),
             ('127', 169),
             ('kg', 130),
             ('2', 8116),
             ('67', 596),
             ('meters', 526),
             ('monster', 298),
             ('looks', 2345),
             ('big', 5696),
             ('enough', 4478),
             ('swallow', 163),
             ('man', 29577),
             ('whole', 1454),
             ("somers'", 162),
             ('sister', 3765),
             ('says', 19676),
             ('yemen', 1592),
             ('apple', 39141),
             ('watch', 27787),
             ('shower', 1323),
             ('proof', 1348),
             ('100', 4837),
             ('000', 13455),
             ('apps', 2225),
             ('launch', 4281),
             ('canadian', 6742),
             ('building', 6338),
             ('multiple', 2904),
             ('shots', 8799),
             ('comcast', 4014),
             ('threatening', 1425),
             ('cut', 2898),
             ('customers', 3026),
             ('use', 7463),
             ('tor', 1667),
             ('web', 2785),
             ('browser', 944),
             ('criminals', 596),
             ('strikes', 5413),
             ('city', 15334),
             ('google', 3664),
             ('buy', 1958),
             ('chunk', 89),
             ('pacific', 771),
             ('shores', 623),
             ('iconic', 679),
             ('redwood', 551),
             ('office', 7972),
             ('park', 2089),
             ('report', 19716),
             ('iraqi', 12209),
             ('social', 7849),
             ('rumors', 6052),
             ('claim', 7669),
             ('leader', 13026),
             ('slain', 399),
             ('there', 31444),
             ('unconfirmed', 1649),
             ('say', 15426),
             ('figment', 29),
             ('your', 11895),
             ('own', 6488),
             ('gross', 173),
             ('imagination', 220),
             ("canada's", 469),
             ("'shot", 28),
             ("dead'", 30),
             ('wounding', 99),
             ('kurds', 2925),
             ('fear', 1940),
             ('isis', 40797),
             ('chemical', 2193),
             ('weapon', 1710),
             ('kobani', 5121),
             ('incident', 5701),
             ('monstrous', 18),
             ('last', 23996),
             ('week', 12491),
             ('hinted', 160),
             ('as', 106611),
             ('ebola', 12490),
             ('fears', 1685),
             ('spread', 3024),
             ('across', 7582),
             ('america', 3742),
             ('today', 6889),
             ('get', 13446),
             ('confirmation', 1988),
             ('daily', 7366),
             ('caller', 228),
             ('passenger', 804),
             ('dulles', 322),
             ('outside', 5354),
             ('washington', 6419),
             ('d', 617),
             ('c', 1649),
             ('apparently', 5395),
             ('taking', 5266),
             ('any', 21470),
             ('chances', 563),
             ('female', 3437),
             ('dressed', 1883),
             ('hazmat', 288),
             ('suit', 696),
             ('complete', 1033),
             ('full', 5185),
             ('body', 5028),
             ('gown', 126),
             ('mask', 463),
             ('gloves', 543),
             ('spotted', 1537),
             ('wednesday', 8421),
             ('waiting', 1505),
             ('flight', 1321),
             ('particularly', 1504),
             ('liked', 316),
             ('jcpenney', 1),
             ('bag', 463),
             ('maybe', 1595),
             ("that's", 3146),
             ('new', 34946),
             ('business', 3243),
             ('bankrupt', 30),
             ('retailer', 166),
             ('side', 2697),
             ('note', 1866),
             ('try', 2507),
             ('halloween', 87),
             ('stores', 1552),
             ('need', 3582),
             ('haz', 1),
             ('mat', 1),
             ('hurry', 9),
             ('begins', 1189),
             ('wearing', 3630),
             ('newser', 156),
             ('-', 11263),
             ('wonder', 1412),
             ('how', 14497),
             ('long', 8715),
             ('quarter', 1306),
             ('pounder', 534),
             ('cheese', 2621),
             ('can', 20106),
             ('two', 23105),
             ('australians', 200),
             ('bought', 1974),
             ('few', 6608),
             ("mcdonald's", 701),
             ('burgers', 212),
             ('friends', 7479),
             ('back', 14268),
             ('1995', 131),
             ('when', 35715),
             ('teens', 850),
             ('never', 8930),
             ('showed', 4435),
             ('up', 37890)
```

```
('up', 37890),
('so', 24608),
("kid's", 56),
('burger', 1932),
('went', 11210),
('uneaten—and', 56),
('stayed', 536),
('way', 10354),
("australia's", 98),
('network', 3018),
('we’re', 1531),
('pretty', 2760),
('sure', 4617),
('oldest', 713),
('men', 5806),
('casey', 283),
('dean', 529),
('holding', 3974),
('onto', 388),
('their', 36284),
('friend', 5208),
('started', 4899),
('joke', 970),
('adds', 1014),
('months', 6687),
('became', 3068),
('years', 15018),
('now', 19995),
('later', 9734),
('same', 8141),
('did', 12541),
('day', 12059),
('perfectly', 737),
('preserved', 264),
('its', 30533),
('original', 2618),
('wrapping', 125),
('his', 113684),
('buying', 840),
('mate', 197),
('eduard', 278),
('nitz', 275),
('even', 17327),
('took', 9447),
('australian', 1172),
('tv', 6504),
('show', 9010),
('project', 2017),
('mold', 103),
('free', 5325),
('specimen', 334),
('9', 6304),
('pair', 1205),
('offered', 2037),
('take', 8262),
('bite', 511),
('charity', 985),
('dissuaded', 56),
("show's", 56),
('hosts', 276),
("they've", 928),
('also', 35924),
('facebook', 8311),
('page', 5341),
('called', 8314),
('year', 27924),
('old', 16253),
('likes', 558),
('than', 20816),
('kanye', 617),
('west', 5817),
('4', 4405),
('044', 56),
('writing', 1660),
("they're", 2333),
('selling', 1475),
('itunes', 547),
('song', 998),
('1', 7747),
('69', 210),
('giving', 1514),
('proceeds', 268),
('beyond', 862),
('blue', 968),
('helps', 605),
('anxiety', 248),
('depression', 118),
('ago', 8419),
('sold', 2698),
('bottle', 560),
('mcjordan', 56),
('sauce', 58),
('10', 9437),
("here's", 1276),
('why', 5523),
('mickey', 56),
("d's", 56),
('food', 5064),
('seemingly', 885),
('decays', 56),
('comfort', 426),
('eating', 2141),
('chinese', 1736),
('woman', 11421),
('26', 2636),
('spends', 777),
('entire', 1972),
('kfc', 2343),
('being', 24418),
('dumped', 892),
('boyfriend', 705),
('suspected', 3925),
('cases', 4633),
('iraq', 18887),
('untrue', 483),
('fighters', 12407),
('contracted', 1609),
("'incorrect'", 60),
('health', 9823),
('ministry', 4143),
('virus', 3738),
('spreading', 751),
('among', 5334),
('militants', 16727),
('jihadists', 2684),
('seeking', 1878),
('cure', 157),
('mosul', 7321),
('hospital', 9456),
("'have", 69),
("ebola'", 69),
('organisation', 1418),
('investigating', 1923),
('lethal', 297),
('disease', 2654),
('cracks', 76),
('five', 5909),
('confirmed', 12177),
('official', 13306),
('confusion', 685),
('swirls', 53),
('details', 4722),
('murky', 114),
('arrest', 2923),
("leader's", 648),
("'wife'", 53),
('outbreak', 783),
('2015', 6739),
('update', 3259),
('may', 19187),
('deadly', 1743),
('sources', 9032),
('allegedly', 6357),
('amazon', 4992),
('com', 7208),
('open', 4788),
('first', 24830),
('physical', 1673),
('manhattan', 1357),
('dj', 119),
('alleged', 6690),
('audio', 8301),
('michael', 11691),
('brown', 14984),
('surfaced', 1826),
('weaponized', 50),
('contract', 1136),
('horror', 333),
('reportedly', 12549),
('matt', 1177),
('taibbi', 1558),
('leave', 3554),
('absence', 2419),
('look', 4925),
("'there", 376),
('limit', 303),
("depravity'", 82),
('islamic', 29507),
('feed', 1410),
('mother', 4753),
('her', 36442),
('son', 8980),
('israeli', 4516),
('fighter', 2841),
('abducted', 2923),
('read', 4165),
('boko', 13586),
('haram', 13226),
('kidnapped', 7576),
('girls', 11969),
('married', 2692),
('truce', 1760),
('truce', 1706),
('does', 5935),
('probably', 4204),
('anyway', 693),
('kurdish', 10159),
('biological', 644),
('warfare', 230),
('horizon', 152),
('soldiers', 4391),
('infected', 979),
('daash', 101),
('transferred', 342),
('translate', 279),
('trick', 271),
('‘eating', 27),
('son’', 80),
('‘have', 57),
('ebola’', 45),
('aren’t', 526),
('attack', 9889),
('britain', 2144),
("'contract", 62),
("virus'", 62),
('grows', 273),
('6', 5212),
('billion', 1434),
('california', 2055),
('deals', 322),
('about', 44194),
('setting', 386),
('straight', 509),
('judicial', 578),
("watch's", 637),
('farrell', 38),
('terrorists', 3896),
('cross', 1439),
('mexican', 2236),
('border', 14796),
('radically', 113),
('redesigned', 119),
('12', 8401),
('inch', 3821),
('macbook', 4281),
('tim', 1654),
('cook', 2669),
('reveals', 1208),
('waterproofing', 174),
('improved', 228),
('getting', 3403),
('dismisses', 52),
('‘unfounded’', 39),
('elderly', 1145),
('arrested', 4605),
('kidnapping', 1414),
('neighbor’s', 201),
('cats', 892),
('making', 6259),
('fur', 576),
('coats', 398),
('denies', 1370),
('5', 5972),
('bird', 704),
('poop', 283),
('vladimir', 565),
('putin', 1984),
('putin', 1084),
('dna', 1385),
('tests', 1439),
('prove', 1181),
('lebanon', 2078),
('chief', 5146),
('al', 27200),
("baghdadi's", 1406),
('ex', 1398),
('wife', 5741),
('daughter', 4584),
('fake', 5472),
('passports', 96),
('officials', 16051),
('refute', 74),
('members', 7027),
('reservations', 155),
('trying', 6108),
('considered', 1383),
('platinum', 752),
('model', 2721),
('former', 8379),
('pga', 2855),
('tour', 4230),
('player', 776),
('tiger', 2224),
('woods', 4052),
('suspended', 1355),
('drug', 3444),
('test', 2068),
('failure', 1275),
('us', 22728),
('journalist', 13100),
('james', 10811),
('foley', 19075),
('hbo', 2477),
('talks', 3731),
('15', 3301),
('month', 10768),
('streaming', 1859),
('service', 11202),
('launching', 509),
('april', 3957),
('investigates', 129),
('blokes', 85),
('dared', 69),
('eat', 1172),
('indian', 1003),
('civil', 2185),
('servant', 235),
('sacked', 431),
("'after", 35),
('24', 3451),
("sickie'", 26),
('contingency', 72),
('plan', 1453),
('developed', 997),
('‘probing’', 57),
('whether', 6564),
('got', 6319),
('posting', 696),
('gun', 2356),
('toting', 1),
('child', 4789),
('online', 7774),
('supporters', 1056),
('announced', 6086),
('group’s', 1434),
('youngest', 933),
('died', 6875),
('combat', 795),
('twitter', 12810),
('accounts', 2059),
('linked', 2572),
('sham', 212),
('claimed', 9075),
('“got', 4),
('martyred”', 1),
('father', 7922),
('while', 21588),
('fighting', 5674),
('terrorist', 5252),
('group', 26105),
('syria', 23848),
('posted', 9326),
('smiling', 916),
('boy', 5258),
('military', 12279),
('fatigues', 525),
('weapons', 6851),
('times', 10074),
('almost', 4251),
('british', 11750),
('roughly', 1226),
('emerged', 2075),
('june', 2882),
('charlie', 1073),
('cooper', 548),
('researcher', 541),
('monitors', 548),
('london', 6999),
('based', 7769),
('quilliam', 241),
('counter', 1884),
('extremism', 346),
('think', 8434),
('tank', 977),
('past', 7059),
('mr', 12311),
('noticed', 1012),
('hashtag', 746),
('“shibal', 1),
('albaghdadi”', 1),
('–', 23438),
('translates', 348),
('“the', 6982),
('cub', 398),
('baghdadi”', 1),
('commonly', 688),
('refer', 237),
('themselves', 2625),
('lions', 60),
('cubs', 35),
('abu', 4841),
('bakr', 2885),
('baghdadi', 5477),
('isis’s', 176),
('self', 1206),
('proclaimed', 665),
('caliph', 219),
('ubaidah', 471),
('martyred', 89),
('airstrikes', 7107),
('weeks', 6138),
('http', 1733),
('t', 2471),
('co', 4877),
('n1puexi7ir', 1),
('istimes1', 1),
('october', 3460),
('07', 71),
('tweeting', 216),
('young', 5924),
('dead', 8148),
('sept', 2534),
('often', 3250),
('produces', 276),
('fabricated', 367),
('make', 10574),
('seem', 2066),
('“more', 104),
('brutal', 2420),
('”', 70057),
('appear', 2684),
('directly', 1245),
('account', 4978),
('u', 25798),
('s', 25838),
('strike', 2973),
('though', 8395),
('identified', 5785),
('“it', 2830),
('seems', 3610),
('like', 20908),
('legitimate', 280),
('thing', 4471),
('telephone', 390),
('interview', 4361),
('thursday', 6111),
('adding', 2769),
('known', 10850),
('“i', 6581),
('would', 36303),
('vouch', 35),
('united', 7015),
('nations', 1397),
('children', 5206),
('conduct', 482),
('patrols', 106),
('guard', 2875),
('prisoners', 988),
('carry', 1122),
('forced', 2482),
('give', 4136),
('blood', 2614),
('injured', 2298),
('listed', 675),
('un', 8946),
('13', 3708),
('frequently', 671),
('used', 9232),
('propaganda', 1421)
```

```
('propaganda', 1421),
('showing', 4268),
('“uniform', 1),
('parading', 35),
('alongside', 2098),
('adults', 113),
('“everyone', 31),
('nato', 286),
('headquarters', 1552),
('worried', 889),
('shelly', 45),
('whitman', 35),
('executive', 1899),
('director', 5264),
('halifax', 35),
('romeo', 63),
('dallaire', 35),
('initiative', 109),
('“if', 1363),
('send', 898),
('boots', 277),
('ground', 4268),
('going', 11477),
('face', 5097),
('she', 40102),
('currently', 3923),
('six', 9209),
('mission', 1532),
('limited', 1353),
('ban', 509),
('deploying', 192),
('troops', 2196),
('‘the', 847),
("baghdadi'", 173),
('jihadist', 2678),
('‘got', 16),
('martyred’', 16),
('least', 10421),
('25', 3133),
('insurgents', 1444),
('clashes', 1013),
('between', 8036),
('islamist', 3513),
('northeast', 793),
('nigeria', 4675),
('civilians', 2170),
('elsewhere', 938),
('region', 3010),
('monday', 9791),
('ceasefire', 4282),
('agreement', 1523),
('nigerian', 6143),
('expected', 4508),
('lead', 2376),
('liberation', 154),
('200', 3834),
('schoolgirls', 3766),
('due', 4484),
('continue', 3385),
('neighbouring', 592),
('chad', 2070),
('attacks', 5870),
('over', 26831),
('blamed', 898),
('security', 16092),
('several', 9704),
('dozen', 1228),
('since', 16328),
('announcement', 2063),
('spokesman', 7343),
('work', 7950),
('criminal', 2250),
('gangs', 572),
('lawless', 181),
('army', 4537),
('officer', 8096),
('requested', 556),
('anonymity', 1663),
('tried', 2279),
('enter', 1093),
('town', 9519),
('damboa', 606),
('late', 5740),
('through', 13213),
('alagarno', 124),
('hideout', 204),
('fought', 1108),
('them', 22252),
('“our', 250),
('gunned', 294),
('entered', 1076),
('unleashed', 192),
('terror', 5795),
('picking', 607),
('ruins', 194),
('armoured', 523),
('vehicle', 3174),
('some', 26496),
('arms', 3059),
('recovered', 954),
('garrison', 102),
('cameroon', 659),
('fierce', 272),
('forces', 10911),
('july', 2596),
('driven', 660),
('out', 34743),
('offensive', 576),
('member', 3173),
...])
```

```
tokenizer.word_docs
```

```
defaultdict(int,
            {'see': 9324,
             'anyone': 4118,
             'sitting': 1162,
             'related': 3050,
             'told': 20533,
             'from': 36033,
             'managua': 976,
             'i': 15632,
             'which': 23489,
             'and': 47088,
             'after': 26886,
             'into': 16499,
             'asteroid': 933,
             'abc': 379,
             'the': 49138,
             'crater': 1259,
             'night': 5767,
             'a': 48457,
             'volcanologist': 332,
             'that': 45001,
             'one': 24201,
             'airport': 2803,
             'national': 5819,
             'mysterious': 845,
             "city's": 496,
             'or': 21113,
             'humberto': 440,
             'radius': 198,
             'scientists': 1311,
             'come': 7282,
             'sudden': 942,
             'saturday': 3247,
             'not': 33664,
             'if': 18066,
             'university': 2064,
             'santamaria': 373,
             'adviser': 511,
             'territorial': 515,
             'reports': 15417,
             'disintegrated': 485,
             'thought': 5841,
             'wave': 857,
             'very': 9068,
             'what': 17698,
             'saw': 3833,
             'force': 3979,
             'at': 35182,
             'heard': 6480,
             'blast': 1012,
             'war': 6109,
             'could': 16293,
             'anything': 2766,
             'wooded': 451,
             'off': 11626,
             'he': 32366,
             'my': 8592,
             'soldier': 2177,
             'help': 5638,
             'locked': 179,
             'visit': 1691,
             'government': 11200,
             'institute': 1636,
             'left': 9077,
             'this': 30105,
             'allowed': 1373,
             'press': 6732,
             'by': 36400,
             'skimmed': 36,
             'passing': 1135,
             'international': 5105,
             'autonomous': 517,
             'strange': 1397,
             'with': 39819,
             'boom': 640,
             'of': 47791,
             'close': 4250,
             'center': 3308,
             'it': 38394,
             'planet': 452,
             'all': 19686,
             'gunfire': 1520,
             'has': 38128,
             'earth': 1507,
             'air': 6402,
             'nothing': 3622,
             'had': 29167,
             'rock': 998,
             'were': 22462,
             'house': 5738,
             'capital': 4597,
             'loud': 1275,
             'media': 13538,
             'is': 42808,
             'nicaragua': 1059,
             'sized': 573,
             '16': 1765,
             'measured': 202,
             'garcia': 297,
             'crashed': 207,
             'rosario': 491,
             'hearing': 1967,
             'state': 15540,
             'erupts': 93,
             'meteorite': 1674,
             'ask': 2105,
             'forecast': 350,
             'wilfried': 394,
             'determined': 1234,
             'pass': 1100,
             'near': 8075,
             'something': 6770,
             'was': 39917,
             'committee': 1743,
             'then': 13584,
             'expansive': 432,
             'down': 9299,
             'depth': 335,
             'an': 35549,
             'sunday': 3846,
             'clear': 4268,
             'deep': 1078,
             'have': 36699,
             'more': 22466,
             'overnight': 749,
             'journalists': 1957,
             'said': 33934,
             'event': 3151,
             'jorge': 455,
             "nicaragua's": 472,
             'we': 20146,
             'happened': 4207,
             'spokeswoman': 2238,
             'they': 26784,
             'news': 16743,
             'still': 9786,
             'relatively': 703,
             '39': 1045,
             'who': 29945,
             'on': 44065,
             'area': 5279,
             'rc': 532,
             'streak': 247,
             'residents': 2704,
             'because': 13081,
             'ice': 363,
             'understanding': 432,
             'parliament': 2257,
             'study': 966,
             'experts': 2686,
             '60': 1354,
             'no': 20680,
             'be': 35181,
             "didn't": 3140,
             'studies': 1092,
             'memorial': 1708,
             'felt': 2762,
             'bomb': 1111,
             'strauch': 422,
             'light': 1609,
             'site': 5723,
             'nicaraguan': 1050,
             'local': 5694,
             'feet': 1534,
             'formed': 638,
             '2014': 6265,
             "managua's": 406,
             'only': 11694,
             'astronomy': 252,
             'murillo': 661,
             'sky': 899,
             'small': 3603,
             'associated': 4035,
             'appears': 5866,
             'reported': 14336,
             'will': 20228,
             'base': 1128,
             'shot': 5420,
             'large': 3917,
             'but': 32674,
             'to': 48368,
             'foot': 1836,
             'porch': 290,
             'diameter': 71,
             'weekend': 3043,
             'saballos': 346,
             'photo': 5529,
             'in': 48095,
             'buried': 1268,
             '‘spider': 30,
             'skin': 2027,
             'skin': 1017,
             'days': 7670,
             'under': 9985,
             'for': 41170,
             'dubbed': 593,
             'man’': 49,
             'tourist': 122,
             'spider': 1551,
             'burrows': 246,
             'somers': 514,
             "yemen'": 70,
             'attempt': 1701,
             'failed': 1770,
             'rescue': 1339,
             "'killed": 124,
             'luke': 639,
             'ottawa': 1803,
             'breaking': 1516,
             'stone': 1011,
             'caught': 3473,
             'reeled': 34,
             'italy': 462,
             'biggest': 992,
             'catfish': 281,
             'line': 2468,
             '8ft': 35,
             '9in': 35,
             'weighing': 313,
             'rod': 209,
             '19': 1832,
             'ever': 4267,
             'giant': 622,
             '20': 3804,
             '40': 3905,
             'minute': 1605,
             'boat': 492,
             'enormous': 589,
             'fishing': 217,
             'battle': 1622,
             'italian': 643,
             'record': 2038,
             'wels': 132,
             'catches': 307,
             'huge': 2021,
             'spice': 1032,
             'pumpkin': 1113,
             'coming': 4170,
             'store': 2776,
             'condom': 798,
             'you': 14928,
             'killed': 9464,
             'gunman': 1316,
             'shooting': 4986,
             'other': 15403,
             'hunt': 544,
             'shooters': 91,
             'hill': 1554,
             'canada': 1300,
             'likely': 5033,
             'surreal': 47,
             'jaw': 37,
             'catch': 763,
             'it’s': 7040,
             'fisherman’s': 24,
             'photos': 2008,
              'photos': 2998,
             'wondering': 387,
             'dropping': 252,
             'people': 18078,
             'real': 6130,
             'lands': 138,
             'hooked': 355,
             'world': 9737,
             'fisherman': 251,
             'source': 5761,
             'wants': 1847,
             'brokaw': 154,
             'williams': 705,
             'fired': 3072,
             'brian': 791,
             'tom': 804,
             'away': 5784,
             'nation’s': 352,
             'been': 32843,
             'canada’s': 144,
             'steps': 718,
             'just': 16172,
             'makes': 2717,
             'set': 5734,
             'pound': 501,
             '280': 128,
             'robots': 348,
             'patrolling': 250,
             'debunked': 599,
             'rumor': 2589,
             'style': 905,
             'robocop': 123,
             'campus': 804,
             "microsoft's": 300,
             'are': 29937,
             'kg': 128,
             'meters': 449,
             '127': 167,
             '67': 434,
             'po': 377,
             '2': 5079,
             'swallow': 162,
             'monster': 277,
             'enough': 3717,
             'man': 13972,
             'big': 4251,
             'whole': 1319,
             'looks': 1931,
             'says': 11999,
             'yemen': 608,
             "somers'": 119,
             'sister': 2563,
             '100': 3052,
             'apple': 6553,
             'proof': 927,
             'apps': 1063,
             'shower': 575,
             'launch': 2863,
             'watch': 7985,
             '000': 7298,
             'canadian': 2540,
             'building': 4013,
             'shots': 2901,
             'multiple': 2734,
             'tor': 346,
             'cut': 2421,
             'comcast': 623,
             'criminals': 543,
             'web': 1992,
             'customers': 1493,
             'browser': 234,
             'use': 4996,
             'threatening': 1251,
             'strikes': 3218,
             'city': 8791,
             'iconic': 677,
             'pacific': 478,
             'redwood': 223,
             'shores': 330,
             'google': 1167,
             'buy': 1263,
             'park': 1587,
             'chunk': 88,
             'office': 4146,
             'report': 12940,
             'social': 6267,
             'leader': 7358,
             'claim': 5709,
             'slain': 303,
             'rumors': 4323,
             'iraqi': 5078,
             'say': 11461,
             'unconfirmed': 1565,
             'there': 20327,
             'imagination': 220,
             'own': 4933,
             'gross': 107,
             'figment': 29,
             'your': 6858,
             "canada's": 253,
             "dead'": 30,
             'wounding': 98,
             "'shot": 28,
             'chemical': 586,
             'weapon': 1207,
             'isis': 13410,
             'kurds': 1842,
             'fear': 1725,
             'kobani': 1794,
             'incident': 2800,
             'monstrous': 18,
             'halloween': 73,
             'mask': 392,
             'as': 35488,
             'wearing': 2535,
             'particularly': 1487,
             'last': 16196,
             'fears': 1386,
             'jcpenney': 1,
             'mat': 1,
             'dulles': 225,
             'suit': 632,
             'haz': 1,
             'body': 4075,
             'america': 3310,
             'c': 1393,
             'd': 566,
             'bag': 362,
             'passenger': 517,
             'maybe': 1395,
             'bankrupt': 30,
             'try': 2230,
             'complete': 1031,
             'dressed': 1587,
             'across': 5722,
             'spotted': 1335,
             'caller': 227,
             'hazmat': 260,
             'waiting': 1457,
             'gloves': 468,
             'new': 17168,
             'note': 1434,
             'outside': 4044,
             'wednesday': 4688,
             'ebola': 3586,
             'liked': 316,
             'washington': 4435,
             'apparently': 4555,
             'need': 3110,
             "that's": 2507,
             'retailer': 166,
             'stores': 591,
             'side': 2223,
             'flight': 1142,
             'begins': 1129,
             'confirmation': 1811,
             'week': 9761,
             'female': 2438,
             'today': 5269,
             'taking': 4833,
             'full': 4483,
             'any': 14017,
             'gown': 126,
             'hurry': 9,
             'spread': 2611,
             'chances': 541,
             'business': 2535,
             'daily': 5097,
             'get': 10023,
             'hinted': 115,
             'dumped': 692,
             'tv': 3899,
             'spends': 688,
             'did': 9642,
             'how': 9993,
             'few': 5611,
             'page': 3798,
             'called': 6988,
             'dissuaded': 56,
             'went': 8608,
             'its': 16303,
             'depression': 118,
             'food': 3456,
             'cheese': 1557,
             'uneaten—and': 56,
             'entire': 1678,
             '69': 210,
             'burgers': 167,
             'hosts': 268,
             'mate': 129,
             'than': 14231,
             'west': 4252
```

```
'west': 4252,
'their': 19153,
'his': 29159,
'also': 22200,
'network': 2509,
'men': 4303,
'way': 8379,
'woman': 6642,
'mcjordan': 56,
'pounder': 285,
'specimen': 245,
'wonder': 1076,
'writing': 1450,
'seemingly': 789,
'newser': 106,
'took': 7679,
"kid's": 56,
'quarter': 857,
'long': 7168,
'burger': 327,
'itunes': 308,
'song': 736,
'holding': 3372,
'joke': 918,
'project': 1600,
'stayed': 446,
'years': 9878,
'helps': 407,
'bought': 1317,
'–': 6174,
'why': 4109,
'buying': 761,
'wrapping': 125,
'later': 8023,
'decays': 56,
'likes': 414,
'boyfriend': 601,
'back': 10393,
'chinese': 848,
'charity': 686,
'day': 8624,
'take': 6998,
'dean': 220,
'now': 14702,
'10': 6496,
"show's": 56,
'sauce': 58,
'we’re': 1482,
'4': 3016,
"here's": 1212,
"australia's": 98,
"they're": 1711,
'became': 2417,
'1995': 128,
'preserved': 222,
'show': 6143,
'friend': 3160,
'mold': 103,
'started': 3716,
'two': 14538,
'even': 12272,
'eating': 1680,
'australian': 957,
'giving': 1408,
'australians': 144,
'perfectly': 737,
'pair': 1007,
'so': 15876,
"mcdonald's": 336,
'teens': 699,
'9': 4718,
'year': 17556,
'selling': 1320,
'being': 17176,
'oldest': 560,
'anxiety': 248,
'1': 5510,
'can': 14138,
'old': 11936,
'044': 56,
'mickey': 56,
'months': 5403,
'ago': 7133,
'adds': 974,
'bite': 498,
'sure': 3802,
'free': 3804,
'kfc': 534,
'original': 2149,
'same': 6239,
'casey': 224,
'kanye': 348,
'facebook': 4953,
'up': 22581,
'onto': 346,
'friends': 4654,
'never': 6394,
'nitz': 212,
'pretty': 2186,
"d's": 56,
'eduard': 220,
"they've": 853,
'offered': 1970,
'beyond': 862,
'sold': 2163,
'bottle': 473,
'comfort': 359,
'proceeds': 260,
'when': 20714,
'showed': 3785,
'blue': 721,
'26': 1939,
'iraq': 8557,
'untrue': 472,
'cases': 2437,
'suspected': 2622,
"'incorrect'": 60,
'health': 4959,
'ministry': 2874,
'fighters': 6867,
'contracted': 1383,
'among': 4301,
'seeking': 1600,
'mosul': 3326,
'virus': 1560,
'hospital': 5254,
'cure': 156,
'spreading': 718,
'jihadists': 1929,
'militants': 7594,
'investigating': 1811,
'disease': 1264,
'lethal': 292,
"'have": 69,
"ebola'": 69,
'organisation': 1048,
'official': 8449,
'five': 3812,
'confirmed': 9977,
'cracks': 76,
"'wife'": 53,
'arrest': 1753,
"leader's": 626,
'murky': 113,
'confusion': 648,
'details': 4260,
'swirls': 53,
'deadly': 1461,
'outbreak': 644,
'update': 2713,
'may': 12515,
'2015': 4088,
'sources': 5897,
'allegedly': 5206,
'first': 15259,
'physical': 1046,
'dj': 119,
'open': 3419,
'com': 4903,
'manhattan': 959,
'amazon': 786,
'brown': 3967,
'audio': 3827,
'surfaced': 1700,
'alleged': 5364,
'michael': 6837,
'contract': 806,
'weaponized': 50,
'reportedly': 9664,
'horror': 319,
'taibbi': 191,
'absence': 1594,
'look': 3554,
'matt': 911,
'leave': 2956,
'her': 10453,
'limit': 300,
'son': 4443,
'islamic': 11027,
"depravity'": 82,
"'there": 375,
'mother': 3304,
'feed': 971,
'fighter': 2377,
'israeli': 1311,
'abducted': 1982,
'read': 3751,
'kidnapped': 4687,
'girls': 3870,
'haram': 3168,
'truce': 1264,
'boko': 2998,
'joke': 2998,
'married': 1641,
'probably': 3618,
'anyway': 658,
'does': 5270,
'kurdish': 4066,
'soldiers': 2652,
'biological': 477,
'warfare': 225,
'infected': 648,
'horizon': 146,
'daash': 79,
'transferred': 342,
'translate': 279,
'son’': 80,
'trick': 262,
'‘eating': 27,
'ebola’': 45,
'‘have': 57,
'aren’t': 467,
"virus'": 62,
"'contract": 62,
'attack': 5956,
'britain': 1771,
'grows': 272,
'california': 1726,
'billion': 1067,
'6': 2745,
'deals': 281,
'about': 23720,
'setting': 386,
'straight': 504,
'border': 5596,
'mexican': 1626,
'judicial': 457,
'farrell': 38,
'cross': 1286,
"watch's": 451,
'terrorists': 2468,
'radically': 110,
'macbook': 758,
'redesigned': 112,
'12': 4701,
'inch': 1077,
'reveals': 1194,
'cook': 1040,
'improved': 219,
'waterproofing': 173,
'tim': 1239,
'getting': 2904,
'‘unfounded’': 39,
'dismisses': 52,
'neighbor’s': 125,
'arrested': 2790,
'making': 5614,
'cats': 195,
'fur': 210,
'elderly': 934,
'coats': 164,
'kidnapping': 1104,
'denies': 1357,
'5': 4361,
'poop': 279,
'bird': 522,
'putin': 409,
'vladimir': 449,
"baghdadi's": 652,
'chief': 3726,
'ex': 1034,
'passports': 96,
'wife': 3141,
'tests': 1022,
'al': 8254,
'fake': 2997,
'lebanon': 998,
'prove': 1102,
'daughter': 1758,
'dna': 744,
'members': 5571,
'refute': 74,
'officials': 9621,
'considered': 1375,
'model': 1562,
'platinum': 372,
'trying': 4887,
'reservations': 153,
'woods': 1181,
'suspended': 929,
'drug': 2227,
'failure': 1085,
'test': 1657,
'former': 5802,
'pga': 623,
'tour': 1402,
'player': 540,
'tiger': 889,
'journalist': 6746,
'us': 12068,
'foley': 6179,
'james': 6699,
'streaming': 723,
'launching': 495,
'service': 4861,
'talks': 2340,
'hbo': 313,
'month': 8404,
'april': 3043,
'15': 2772,
'investigates': 129,
'dared': 69,
'blokes': 84,
'eat': 1067,
'indian': 875,
'24': 3020,
'servant': 166,
'civil': 1743,
"'after": 35,
"sickie'": 26,
'sacked': 363,
'contingency': 71,
'plan': 1309,
'developed': 979,
'got': 5351,
'whether': 5429,
'‘probing’': 57,
's': 9945,
'london': 3601,
'nato': 283,
'“uniform': 1,
'interview': 3812,
'noticed': 966,
'showing': 3743,
'injured': 1922,
'accounts': 1929,
'listed': 567,
'“i': 4507,
'“more': 104,
'identified': 4168,
'shelly': 45,
'halifax': 35,
'children': 3171,
'vouch': 35,
'emerged': 1658,
'syria': 8771,
'supporters': 831,
'commonly': 669,
'directly': 1178,
'currently': 3495,
'cooper': 354,
'counter': 1410,
"baghdadi'": 168,
'adults': 111,
'parading': 35,
'blood': 2260,
'past': 5917,
'lions': 60,
'father': 3791,
'legitimate': 280,
'conduct': 454,
'posted': 6118,
'often': 2699,
'”': 17428,
'weapons': 3128,
'albaghdadi”': 1,
'known': 8329,
'worried': 845,
'–': 9726,
'ubaidah': 180,
'martyred': 89,
'cub': 264,
'forced': 2116,
'dallaire': 35,
'un': 3401,
'headquarters': 1493,
'troops': 1806,
'roughly': 1123,
'tank': 836,
'“it': 2417,
'smiling': 446,
'tweeting': 216,
'thing': 3458,
'martyred”': 1,
'carry': 1101,
'while': 15035,
'researcher': 541,
'child': 2765,
'proclaimed': 627,
'group': 11734,
'caliph': 160,
'martyred’': 16,
'october': 2609,
'“got': 4,
'baghdadi”': 1,
'guard': 1859,
'alongside': 1602,
'http': 1400,
'istimes1': 1,
'hashtag': 544,
'mission': 1196,
'sept': 1534,
'online': 5967,
'youngest': 538,
'propaganda': 1182,
'almost': 3729,
'based': 6809,
'weeks': 4595,
'like': 14704,
'patrols': 106,
'appear': 2234,
'abu': 2532,
'toting': 1,
'u': 9763,
'director': 4072,
'jihadist': 2165,
'times': 7466,
'initiative': 87,
'self': 1160,
'british': 4673,
'fighting': 3776,
'whitman': 35,
'romeo': 63,
'limited': 1256,
'nations': 1361,
'group’s': 1244,
'quilliam': 241,
'terrorist': 3502,
'june': 2506,
'“everyone': 31,
'“if': 1315,
'baghdadi': 1831,
'twitter': 8254,
'sham': 179,
'07': 71,
'make': 8338,
'young': 4420,
'fatigues': 380,
'monitors': 546,
'frequently': 636,
'gun': 1634,
'combat': 738,
'adding': 2476,
'deploying': 192,
'seem': 1797,
'thursday': 4133,
'ban': 390,
'brutal': 2078,
'six': 6445,
'she': 11696,
'bakr': 1659,
'united': 4931,
'themselves': 2426,
'account': 3452,
't': 2064,
'extremism': 346,
'‘got': 16,
'“the': 5273
```

```
'the': 5273,
'charlie': 791,
'mr': 3936,
'strike': 1657,
'prisoners': 806,
'used': 6592,
'“shibal': 1,
'claimed': 6585,
'seems': 3261,
'died': 4585,
'boots': 273,
'telephone': 375,
'refer': 202,
'co': 3770,
'13': 2540,
'linked': 2094,
'executive': 1494,
'translates': 348,
'send': 876,
'dead': 4862,
'give': 3644,
'boy': 2630,
'think': 6380,
'would': 19311,
'produces': 276,
'isis’s': 174,
'fabricated': 336,
'though': 6798,
'airstrikes': 3453,
'cubs': 35,
'military': 6247,
'going': 8181,
'face': 3489,
'posting': 622,
'ground': 3864,
'announced': 4680,
'‘the': 780,
'n1puexi7ir': 1,
'ceasefire': 1835,
'picking': 607,
'work': 5177,
'tada': 102,
'neighbouring': 565,
'nigerian': 1841,
'entered': 1049,
'tried': 1912,
'scouting': 231,
'beheads': 1669,
'mohammed': 2081,
'continue': 2751,
'far': 4338,
'american': 9897,
'chad': 1188,
'over': 17743,
'ruins': 194,
'unleashed': 192,
'liberation': 154,
'hilly': 102,
'maiduguri': 281,
'morning': 3706,
'since': 10991,
'midnight': 603,
'monday': 6262,
'lead': 2207,
'mountain': 1077,
'slaughtered': 197,
'anonymity': 1541,
'fought': 1074,
'“our': 250,
'gangs': 552,
'cameroon': 382,
'“two': 110,
'damboa': 139,
'member': 2940,
'200': 3035,
'alagarno': 124,
'“they': 1382,
'clashes': 708,
'arms': 1741,
'between': 6446,
'civilians': 1833,
'joint': 1304,
'phone': 2633,
'engaged': 420,
'terror': 3878,
'least': 7650,
'armoured': 378,
'elsewhere': 844,
'gunned': 279,
'reuters': 2853,
'late': 4952,
'andrew': 389,
'region': 2275,
'lasted': 321,
'blamed': 710,
'recovered': 813,
'spokesman': 5944,
'northeast': 670,
'resident': 1329,
'nigeria': 1945,
'insurgents': 755,
'task': 618,
'civilian': 774,
'announcement': 1842,
'till': 148,
'several': 7614,
'schoolgirls': 1928,
'relatives': 835,
'town': 4995,
'attacks': 3503,
...})
```

```
word_index = tokenizer.word_index
```

```
tokenizer.document_count
```

```
49972
```

fit_on_texts() gives the following attributes in the output as given here.

- **word_counts:** dictionary mapping words (str) to the number of times they appeared on during fit. Only s

- **word_docs:** dictionary mapping words (str) to the number of documents/texts they appeared on during called.
- **word_index:** dictionary mapping words (str) to their rank/index (int). Only set after fit_on_texts was calle
- **document_count:** int. Number of documents (texts/sequences) the tokenizer was trained on. Only set a

Now, tokenize the sentences using nltk sent_tokenize() and encode the senteces with tl above `t.word_index`

Initialise 2 lists with names `texts` and `articles`.

```
texts = [] to store text of article as it is.

articles = [] split the above text into a list of sentences.
```

```
texts = []
articles = []
```

```
from nltk.tokenize import sent_tokenize

# Method 1
texts = dataset['articleBody']
for items in texts.iteritems():
    articles.append(sent_tokenize(items[1]))
```

Check 2:

first element of texts and articles should be as given below.

```
texts[0]
```

```
'A small meteorite crashed into a wooded area in Nicaragua\'s capital of Managua overnight, the
```

```
articles[0]
```

```
['A small meteorite crashed into a wooded area in Nicaragua's capital of Managua overnight, the
 "Residents reported hearing a mysterious boom that left a 16-foot deep crater near the city's
 'Government spokeswoman Rosario Murillo said a committee formed by the government to study the
 'House-sized asteroid 2014 RC, which measured 60 feet in diameter, skimmed the Earth this week
 'Murillo said Nicaragua will ask international experts to help local scientists in understandi
 'The crater left by the meteorite had a radius of 39 feet and a depth of 16 feet,  said Humber
 'He said it is still not clear if the meteorite disintegrated or was buried.',
 'Humberto Garcia, of the Astronomy Center at the National Autonomous University of Nicaragua,
 '"We have to study it more because it could be ice or rock," he said.',
 'Wilfried Strauch, an adviser to the Institute of Territorial Studies, said it was "very stran
 'We have to ask if anyone has a photo or something."',
 "Local residents reported hearing a loud boom Saturday night, but said they didn't see anythin
 '"I was sitting on my porch and I saw nothing, then all of a sudden I heard a large blast.',
 'We thought it was a bomb because we felt an expansive wave," Jorge Santamaria told The Associ
 "The site of the crater is near Managua's international airport and an air force base.",
 'Only journalists from state media were allowed to visit it.']
```

Now iterate through each article and each sentence to encode the t.word_index [5 marks]

Here, to get words from sentence you can use `text_to_word_sequence` from keras preprocessing text.

1. Import text_to_word_sequence
2. Initialize a variable of shape (no.of articles, MAX_SENTS, MAX_SENT_LENGTH) with name `data` with ze np.zeros to initialize with all zeros)and then update it while iterating through the words and sentences i

```
from keras.preprocessing.text import text_to_word_sequence
import numpy as np
```

```
data = np.zeros((len(articles), MAX_SENTS, MAX_SENT_LENGTH)).astype(int)
data.shape
```

```
(49972, 20, 20)
```

```
data[0, :, :]
```

```
array([[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
       [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
       [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
       [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
       [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
       [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
       [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
       [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
       [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
       [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
       [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
       [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
       [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
       [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
       [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
       [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
       [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
       [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
       [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
       [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]])
```

```
for pos1, item in enumerate(articles):
    for pos2, sent in enumerate(item):
        if(sent is not None) & (pos2 < MAX_SENTS):
            temp = text_to_word_sequence(sent)
            for pos3, word in enumerate(temp):
                if (pos3 < MAX_SENT_LENGTH):
                    data[pos1][pos2][pos3] = word_index.get(word, 0)
```

```
data[0, :, :]
```

```
array([[    3,   485,   433,  7204,    81,     3,  3732,   331,     5,
         3888,   350,     4,  1432,  2956,     1,    89,    12,   464,
            0,     0],
       [  757,    95,  1045,     3,  2675,  1750,     7,   188,     3,
         1217,  1074,  2026,   698,   158,     1,  3029,   449,     1,
          555,   243],
       [   89,  1065,  4111,  2345,    12,     3,  1092,  3300,    19,
            1,    89,     2,  1791,     1,   529,  2005,    15,     9,
            3,  3107],
       [  186,  3639,   971,   202,  2553,    43,  6770,  1719,  1250,
            5, 13306, 17921,     1,   776,    31,   738,  3986,    67,
           85,     0],
       [ 2345,    12,  1584,    38,  1094,   351,   777,     2,   367,
          260,  1775,     5,  4447,    70,   494,     0,     0,     0,
            0,     0],
       [    1,   698,   188,    19,     1,   433,    32,     3,  7411,
            4,  2256,  1250,     6,     3,  5266,     4,  1217,  1250,
           12,  3359],
       [   13,    12,    15,     8,   148,    25,   541,    64,     1,
          433,  3726,    41,     9,  1848,     0,     0,     0,     0,
            0,     0],
       [ 3359,  5729,     4,     1,  5869,   613,    21,     1,   308,
         3432,   796,     4,  1584,    12,     1,   433,    69,    23,
          785,     2],
       [   37,    17,     2,  1791,    15,    52,   120,    15,    69,
           23,  4916,    41,  1960,    13,    12,     0,     0,     0,
            0,     0],
       [ 4733,  3334,    24,  3965,     2,     1,  1314,     4,  3067,
         1651,    12,    15,     9,   195,  1423,     7,    58,    40,
           95,     3],
       [   37,    17,     2,  1094,    64,   509,    20,     3,   250,
           41,   264,     0,     0,     0,     0,     0,     0,     0,
            0,     0],
       [  260,   757,    95,  1045,     3,  1804,  1750,   530,   275,
           28,    12,    33,   700,   163,   891,  1423,     5,     1,
         2078,     0],
       [   35,     9,  2057,    10,   116,  5820,     6,    35,   574,
          655,   104,    59,     4,     3,  2407,    35,   240,     3,
          511,  1910],
       [   37,   340,    15,     9,     3,  2079,   120,    37,   880,
           24,  4448,  2581,  4315,  4917,    55,     1,   555,   243,
            0,     0],
       [    1,   255,     4,     1,   698,     8,   158,  3957,   351,
          449,     6,    24,   154,   465,  1926,     0,     0,     0,
            0,     0],
       [  126,   921,    22,    47,   100,    36,  1832,     2,  1212,
           15,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0],
       [    0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0],
       [    0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0],
       [    0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0],
       [    0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0]])
```

Check 3:

Accessing first element in data should give something like given below.

```
data[0, :, :]
```

```
array([[    3,   487,   474,  7113,    79,     3,  3687,   325,     5,
         4200,   361,     4,  1525,  2913,     1,    89,    12,   451,
            0,     0],
       [  743,    96,  1044,     3,  2814,  1759,     7,   186,     3,
         1219,  1070,  1987,   736,   154,     1,  2990,   458,     1,
          543,   232],
       [   89,  1052,  4057,  2314,    12,     3,  1073,  3248,    19,
            1,    89,     2,  1751,     1,   518,  1980,    15,     9,
            3,  2879],
       [  182,  3691,   976,   196,  2515,    42,  6688,  1691,  1227,
            5, 13011, 17379,     1,   762,    30,   722,  3931,    66,
           87,     0],
       [ 2314,    12,  1882,    38,  1076,   346,   793,     2,   356,
          261,  1782,     5,  4396,    67,   486,     0,     0,     0,
            0,     0],
       [    1,   736,   186,    19,     1,   474,    32,     3,  7307,
            4,  2122,  1227,     6,     3,  5195,     4,  1219,  1227,
           12,  3308],
       [   13,    12,    15,     8,   143,    25,   531,    63,     1,
          474,  3679,    41,     9,  1825,     0,     0,     0,     0,
            0,     0],
       [ 3308,  5643,     4,     1,  5788,   620,    22,     1,   302,
         3125,   786,     4,  1882,    12,     1,   474,    70,    23,
          801,     2],
       [   35,    17,     2,  1751,    15,    54,   119,    15,    70,
           23,  4850,    41,  1885,    13,    12,     0,     0,     0,
            0,     0],
       [ 4664,  3279,    24,  3915,     2,     1,  1298,     4,  3028,
         1630,    12,    15,     9,   187,  1423,     7,    56,    40,
           96,     3],
       [   35,    17,     2,  1076,    63,   497,    20,     3,   252,
           41,   260,     0,     0,     0,     0,     0,     0,     0,
            0,     0],
       [  261,   743,    96,  1044,     3,  1765,  1759,   520,   273,
           29,    12,    33,   702,   160,   818,  1423,     5,     1,
         2068,     0],
       [   34,     9,  2035,    10,   112,  5741,     6,    34,   562,
          644,   104,    57,     4,     3,  2382,    34,   238,     3,
          504,  1922],
       [   35,   341,    15,     9,     3,  2053,   119,    35,   872,
           24,  4397,  2541,  4258,  4851,    55,     1,   543,   232,
            0,     0],
       [    1,   254,     4,     1,   736,     8,   154,  4116,   346,
          458,     6,    24,   152,   460,  1908,     0,     0,     0,
            0,     0],
       [  124,   896,    21,    48,   102,    37,  1803,     2,  1195,
           15,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0],
       [    0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0],
       [    0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0],
       [    0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0],
       [    0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0]], dtype=int32)
```

Repeat the same process for the `Headings` as well. Use variables w `texts_heading` and `articles_heading` accordingly. [5 marks]

```
texts_headings = []
articles_headings = []

texts_headings = dataset['Headline']
for items in texts_headings.iteritems():
    articles_headings.append(sent_tokenize(items[1]))
```

```
data_heading = np.zeros((len(articles_headings), MAX_SENTS, MAX_SENT_LENGTH)).astype(int)
data_heading.shape
```

```
(49972, 20, 20)
```

```
for pos1, item in enumerate(articles_headings):
    for pos2, sent in enumerate(item):
        if(sent is not None) & (pos2 < MAX_SENTS):
            temp = text_to_word_sequence(sent)
            for pos3, word in enumerate(temp):
                if (pos3 < MAX_SENT_LENGTH):
                    data_heading[pos1][pos2][pos3] = word_index.get(word, 0)
```

```
data_heading[4, :, :]
```

```
array([[ 3030, 12886, 17922,  1080,  5468,  1176,  1971,   450,     5,
         2926,     8,   340,     2,    23,     1,  1841,   520, 18040,
            5,    14],
       [    0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0],
       [    0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0],
       [    0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0],
       [    0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0],
       [    0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0],
       [    0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0],
       [    0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0],
       [    0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0],
       [    0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0],
       [    0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0],
       [    0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0],
       [    0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0],
       [    0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0],
       [    0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0],
       [    0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0],
       [    0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0],
       [    0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0],
       [    0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0],
       [    0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0,     0,     0,     0,     0,     0,     0,     0,
            0,     0]])
```

Now the features are ready, lets make the labels ready for the model to process.

Convert labels into one-hot vectors

You can use get_dummies in pandas to create one-hot vectors.

```
labels = pd.get_dummies(dataset['Stance'])
labels.head()
```

	agree	disagree	discuss	unrelated
0	0	0	0	1
1	0	0	0	1
2	0	0	0	1
3	0	0	0	1
4	0	0	0	1

Check 4:

The shape of data and labels shoould match the given below numbers.

```
print('Shape of data tensor:', data.shape)
print('Shape of label tensor:', labels.shape)
```

```
Shape of data tensor: (49972, 20, 20)
Shape of label tensor: (49972, 4)
```

Shuffle the data

```
## get numbers upto no.of articles
indices = np.arange(data.shape[0])
## shuffle the numbers
np.random.shuffle(indices)
```

```
## shuffle the data
data = data[indices]
data_heading = data_heading[indices]
## shuffle the labels according to data
labels = labels.iloc[indices]
```

Split into train and validation sets. Split the train set 80:20 ratio to get the train and valic

Use the variable names as given below:

x_train, x_val - for body of articles.

x-heading_train, x_heading_val - for heading of articles.

y_train - for training labels.

y_val - for validation labels.

```
from sklearn.model_selection import train_test_split
x_train,x_val,x_heading_train,x_heading_val,y_train,y_val =train_test_split(data,data_heading,labels
```

```
x_train.shape
```

```
(39977, 20, 20)
```

```
x_val.shape
```

```
(9995, 20, 20)
```

```
x_heading_train.shape
```

```
(39977, 20, 20)
```

```
x_heading_val.shape
```

```
(9995, 20, 20)
```

```
y_train.shape
```

```
(39977, 4)
```

```
y_val.shape
```

```
(9995, 4)
```

Check 5:

The shape of x_train, x_val, y_train and y_val should match the below numbers.

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

print(x_val.shape)
print(y_val.shape)
```

```
(39977, 20, 20)
(39977, 4)
(9995, 20, 20)
(9995, 4)
```

Create embedding matrix with the glove embeddings

Run the below code to create embedding_matrix which has all the words and their glove embedding if presen

```
import numpy as np
embeddings_index = dict()
f = open('./glove.6B.100d.txt')
```

```
for line in f:
    values = line.split()
    word = values[0]
    coefs = np.asarray(values[1:], dtype='float32')
    embeddings_index[word] = coefs
f.close()
print('Loaded %s word vectors.' % len(embeddings_index))

# create a weight matrix for words in training docs
embedding_matrix = np.zeros((len(word_index) + 1, 100))


for word, i in tokenizer.word_index.items():
    embedding_vector = embeddings_index.get(word)
    if embedding_vector is not None:
        embedding_matrix[i] = embedding_vector
```

```
Loaded 400000 word vectors.
```

Try the sequential model approach and report the accuracy score. [

Import layers from Keras to build the model

```
from tensorflow import keras
from keras import backend as K
import numpy as np
from keras.layers import LSTM, Dense, Dropout, Embedding, Masking, Bidirectional, GlobalAveragePooli
```

Model

```
model = keras.Sequential()
model.add(keras.layers.Embedding(len(word_index) + 1, 100))
model.add(keras.layers.BatchNormalization())
model.add(keras.layers.GlobalAveragePooling1D())
model.add(keras.layers.Dense(64, activation='relu'))

# Dropout for regularization
model.add(keras.layers.Dropout(0.5))
model.add(keras.layers.Dense(32, activation="relu"))
model.add(keras.layers.Dense(16, activation="relu"))
model.add(keras.layers.Dense(4, activation="sigmoid"))
```

```
model.summary()
```

```
Model: "sequential_5"
_________________________________________________________________
Layer (type)                 Output Shape              Param #   
=================================================================
embedding_5 (Embedding)      (None, None, 100)         3364300   
_________________________________________________________________
batch_normalization_2 (Batch (None, None, 100)         400       
_________________________________________________________________
global_average_pooling1d_4 ( (None, 100)               0         
_________________________________________________________________
dense_12 (Dense)             (None, 64)                6464      
_________________________________________________________________
dropout_3 (Dropout)          (None, 64)                0         
_________________________________________________________________
dense_13 (Dense)             (None, 32)                2080      
_________________________________________________________________
dense_14 (Dense)             (None, 16)                528       
_________________________________________________________________
dense_15 (Dense)             (None, 4)                 68        
=================================================================
Total params: 3,373,840
Trainable params: 3,373,640
Non-trainable params: 200
_________________________________________________________________
```

Compile and fit the model

```
model.compile(optimizer='adam', loss='categorical_crossentropy', metrics=['accuracy'])
```

```
x_train = np.reshape(x_train, (39977, 400))
x_val = np.reshape(x_val, (9995, 400))
```

```
 model.fit(x_train, y_train, validation_data=(x_val, y_val), epochs = 20, batch_size = 80, verbose =
Epoch 1/20
500/500 [==============================] - 24s 48ms/step - loss: 0.7913 - accuracy: 0.7310 - va
Epoch 2/20
500/500 [==============================] - 24s 48ms/step - loss: 0.7865 - accuracy: 0.7310 - va
Epoch 3/20
500/500 [==============================] - 24s 48ms/step - loss: 0.7837 - accuracy: 0.7310 - va
Epoch 4/20
500/500 [==============================] - 24s 48ms/step - loss: 0.7807 - accuracy: 0.7310 - va
Epoch 5/20
500/500 [==============================] - 24s 48ms/step - loss: 0.7783 - accuracy: 0.7310 - va
Epoch 6/20
500/500 [==============================] - 24s 48ms/step - loss: 0.7755 - accuracy: 0.7310 - va
Epoch 7/20
500/500 [==============================] - 24s 48ms/step - loss: 0.7736 - accuracy: 0.7310 - va
Epoch 8/20
500/500 [==============================] - 24s 47ms/step - loss: 0.7726 - accuracy: 0.7310 - va
Epoch 9/20
500/500 [==============================] - 24s 47ms/step - loss: 0.7714 - accuracy: 0.7310 - va
Epoch 10/20
500/500 [==============================] - 24s 47ms/step - loss: 0.7696 - accuracy: 0.7310 - va
Epoch 11/20
500/500 [==============================] - 24s 47ms/step - loss: 0.7692 - accuracy: 0.7310 - va
Epoch 12/20
500/500 [==============================] - 24s 47ms/step - loss: 0.7673 - accuracy: 0.7310 - va
Epoch 13/20
500/500 [==============================] - 24s 47ms/step - loss: 0.7669 - accuracy: 0.7310 - va
Epoch 14/20
500/500 [==============================] - 24s 47ms/step - loss: 0.7660 - accuracy: 0.7310 - va
Epoch 15/20
500/500 [==============================] - 24s 47ms/step - loss: 0.7649 - accuracy: 0.7310 - va
Epoch 16/20
500/500 [==============================] - 24s 47ms/step - loss: 0.7648 - accuracy: 0.7310 - va
Epoch 17/20
500/500 [==============================] - 24s 48ms/step - loss: 0.7644 - accuracy: 0.7310 - va
Epoch 18/20
500/500 [==============================] - 24s 48ms/step - loss: 0.7626 - accuracy: 0.7310 - va
Epoch 19/20
500/500 [==============================] - 24s 48ms/step - loss: 0.7621 - accuracy: 0.7310 - va
Epoch 20/20
500/500 [==============================] - 24s 48ms/step - loss: 0.7618 - accuracy: 0.7310 - va
<tensorflow.python.keras.callbacks.History at 0x7f167e46c9e8>
```

```
# Steps for accuracy
# Get back the encoding values to compare with the predictions for getting accuracy score
i, val = np.where(y_val)

#Get predictions for validation data
pred = model.predict_classes(x_val)
```

```
from sklearn.metrics import accuracy_score

print ('Accuracy of the model is: ', accuracy_score(val, pred)*100)
```

```
Accuracy of the model is:  73.23661830915458
```

Build the same model with attention layers included for better performance

Fit the model and report the accuracy score for the model with attention lay