

ex4-nlp

February 14, 2024

0.0.1 Importing Packages

```
[243]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import re
import nltk
from nltk.corpus import stopwords
from nltk.tokenize import TweetTokenizer
from nltk.stem import PorterStemmer, WordNetLemmatizer
from tensorflow.keras.preprocessing.text import Tokenizer
from tensorflow.keras.preprocessing.sequence import pad_sequences
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Embedding, Conv1D, GlobalMaxPooling1D, Dense, SimpleRNN
from gensim.models.word2vec import Word2Vec
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import *
```

0.0.2 Downloading the resources

```
[10]: nltk.download('stopwords')
nltk.download('wordnet')
```

```
[nltk_data] Downloading package stopwords to C:\Users\Muthu
[nltk_data] Palaniappan M\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[nltk_data] Downloading package wordnet to C:\Users\Muthu Palaniappan
[nltk_data] M\AppData\Roaming\nltk_data...
[nltk_data] Package wordnet is already up-to-date!
```

```
[10]: True
```

0.0.3 Loading Data

```
[2]: data = pd.read_csv("labeled_data.csv")
data.head()
```

```
[2]: Unnamed: 0  count  hate_speech  offensive_language  neither  class  \
0          0      3          0              0          3      2
1          1      3          0              3          0      1
2          2      3          0              3          0      1
3          3      3          0              2          1      1
4          4      6          0              6          0      1
```

```
                                tweet
0  !!! RT @mayasolovely: As a woman you shouldn't...
1  !!!!! RT @mleew17: boy dats cold...tyga dwn ba...
2  !!!!!!! RT @UrKindOfBrand Dawg!!!! RT @80sbaby...
3  !!!!!!!! RT @C_G_Anderson: @viva_based she lo...
4  !!!!!!!!!!!!!!! RT @ShenikaRoberts: The shit you...
```

```
[5]: data = data[['tweet', 'class']]
```

```
[6]: data
```

```
[6]:                                tweet  class
0      !!! RT @mayasolovely: As a woman you shouldn't...      2
1      !!!!! RT @mleew17: boy dats cold...tyga dwn ba...      1
2      !!!!!!! RT @UrKindOfBrand Dawg!!!! RT @80sbaby...      1
3      !!!!!!!! RT @C_G_Anderson: @viva_based she lo...      1
4      !!!!!!!!!!!!!!! RT @ShenikaRoberts: The shit you...      1
...
24778  you's a muthaf***in lie &#8220;@LifeAsKing: @2...      1
24779  you've gone and broke the wrong heart baby, an...      2
24780  young buck wanna eat!!.. dat nigguh like I ain...      1
24781      youu got wild bitches tellin you lies      1
24782  ~~Ruffled | Ntac Eileen Dahlia - Beautiful col...      2
```

```
[24783 rows x 2 columns]
```

```
[11]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 24783 entries, 0 to 24782
Data columns (total 2 columns):
#   Column  Non-Null Count  Dtype
---  -
0   tweet   24783 non-null   object
1   class   24783 non-null   int64
dtypes: int64(1), object(1)
```

memory usage: 387.4+ KB

0.0.4 TweetProcessing

word_tokenize has split #dummysmile as '#' and 'dummysmile', while TweetTokenizer didn't, as '#dummysmile'. TweetTokenizer is built mainly for analyzing tweets.

```
[16]: stop_words = set(stopwords.words('english'))
      len(stop_words)
```

```
[16]: 179
```

```
[17]: lemmatizer = WordNetLemmatizer()
      tokenizer = TweetTokenizer()
```

Tokenize

```
[71]: tokens = tokenizer.tokenize("Can't believe it! @Muthupa just posted this_
      ↪amazing #welcome to snuc https://snu.com #excited")
      tokens
```

```
[71]: ["Can't",
      'believe',
      'it',
      '!',
      '@Muthupa',
      'just',
      'posted',
      'this',
      'amazing',
      '#welcome',
      'to',
      'snuc',
      ' ',
      'https://snu.com',
      '#excited']
```

Removing url's and stopwords

```
[43]: tokens = [re.sub(r'https?:\/\/\S+|www\.\S+', '', token) for token in tokens]
      tokens
```

```
[43]: ["Can't",
      'believe',
      'it',
      '!',
      '@Muthupa',
      'just',
      'posted',
```

```
'this',  
'amazing',  
'#welcome',  
'to',  
'snuc',  
' ',  
'',  
'#excited']
```

```
[44]: tokens = [token for token in tokens if token not in stop_words]  
tokens
```

```
[44]: ["Can't",  
      'believe',  
      '!',  
      '@Muthupa',  
      'posted',  
      'amazing',  
      '#welcome',  
      'snuc',  
      ' ',  
      '',  
      '#excited']
```

Lemmatize

```
[35]: tokens = [lemmatizer.lemmatize(token) for token in tokens]  
tokens
```

```
[35]: ["Can't",  
      'believe',  
      '!',  
      '@Muthupa',  
      'posted',  
      'amazing',  
      '#welcome',  
      'snuc',  
      ' ',  
      '',  
      '#excited']
```

Emoji Mapping

```
[58]: tokens = [emoji.demojize(token) for token in tokens]  
tokens
```

```
[58]: ["Can't",  
      'believe',
```

```
'!',
'@Muthupa',
'posted',
'amazing',
'#welcome',
'snuc',
':star-struck:',
'',
'#excited']
```

Removing Userhandles

```
[67]: tokens = [token for token in tokens if not token.startswith('@')]
tokens
```

```
[67]: ["Can't",
'believe',
'!',
'posted',
'amazing',
'#welcome',
'snuc',
':star-struck:',
'',
'#excited']
```

```
[68]: processed_tweet = ' '.join(tokens)
processed_tweet
```

```
[68]: "Can't believe ! posted amazing #welcome snuc :star-struck: #excited"
```

```
[113]: def preprocess_tweet(tweet):
tokens = tokenizer.tokenize(tweet)
tokens = [token.lower() for token in tokens]
tokens = [re.sub(r'https?://\S+|www\.\S+', '', token) for token in tokens]
tokens = [re.sub(r'\W', ' ', token) for token in tokens if token.isalnum()]
tokens = [token for token in tokens if token not in stop_words]
tokens = [lemmatizer.lemmatize(token) for token in tokens]
tokens = [emoji.demojize(token) for token in tokens]
contractions = {
    "can't": "cannot",
    "won't": "will not",
}
tokens = [contractions[token] if token in contractions else token for token
in tokens]
tokens = [token for token in tokens if not token.startswith('@')]
processed_tweet = ' '.join(tokens)
```

```
return processed_tweet
```

Pre-Processing Function

```
[72]: tweet_pre_process("Can't believe it! @Muthupa just posted this amazing #welcome_
↳to snuc https://snu.com #excited")
```

```
[72]: "can't believe ! posted amazing #welcome snuc :star-struck: #excited"
```

```
[114]: data['pre_process_tweet'] = data['tweet'].apply(preprocess_tweet)
```

```
[118]: data
```

```
[118]:
```

	tweet	class	\
0	!!! RT @mayasolovely: As a woman you shouldn't...	2	
1	!!!! RT @mleew17: boy dats cold...tyga dwn ba...	1	
2	!!!!!! RT @UrKindOfBrand Dawg!!!! RT @80sbaby...	1	
3	!!!!!! RT @C_G_Anderson: @viva_based she lo...	1	
4	!!!!!! RT @ShenikaRoberts: The shit you...	1	
...	
24778	you's a muthaf***in lie “@LifeAsKing: @2...	1	
24779	you've gone and broke the wrong heart baby, an...	2	
24780	young buck wanna eat!!.. dat nigguh like I ain...	1	
24781	youu got wild bitches tellin you lies	1	
24782	~~Ruffled Ntac Eileen Dahlia - Beautiful col...	2	

	pre_process_tweet
0	rt woman complain cleaning house man always ta...
1	rt boy dat cold tyga dwn bad cuffin dat hoe 1s...
2	rt dawg rt 0sbaby4life ever fuck bitch start c...
3	rt look like tranny
4	rt shit hear might true might faker bitch told ya
...	...
24778	muthaf lie right tl trash mine bible scripture...
24779	gone broke wrong heart baby drove redneck crazy
24780	young buck wanna eat dat nigguh like aint fuck...
24781	youu got wild bitch tellin lie
24782	ruffled ntac eileen dahlia beautiful color com...

```
[24782 rows x 3 columns]
```

0.1 Word2Vec - CBOW

```
[212]: cbow = Word2Vec(data['pre_process_tweet'].values.tolist(), vector_size=100,
    ↪window=5, min_count=2, sg=0)
vocab = cbow.wv.index_to_key
```

```
[213]: def get_mean_vector(model, sentence):
    words = [word for word in sentence if word in vocab]
    if len(words) >= 1:
        return np.mean(model.wv[words], axis=0)
    return np.zeros((100,))
```

```
[214]: cbow_array = []
for sentence in data['pre_process_tweet'].values.tolist():
    cbow_array.append(get_mean_vector(cbow, sentence))
```

```
[216]: len(cbow_array)
```

```
[216]: 24782
```

0.2 Word2Vec - Skipgram

```
[219]: sg = Word2Vec(data['pre_process_tweet'].values.tolist(), vector_size=100,
    ↪window=5, min_count=2, sg=1)
vocab = sg.wv.index_to_key
```

```
[220]: def get_mean_vector(model, sentence):
    words = [word for word in sentence if word in vocab]
    if len(words) >= 1:
        return np.mean(model.wv[words], axis=0)
    return np.zeros((100,))
```

```
[221]: sg_array = []
for sentence in data['pre_process_tweet'].values.tolist():
    sg_array.append(get_mean_vector(sg, sentence))
```

```
[223]: len(sg_array)
```

```
[223]: 24782
```

0.3 CNN - Representation

0.3.1 Dataset_Preparation

```
[175]: tweets = data['pre_process_tweet'].values
labels = data['class'].values
```

```
[176]: tweets
```

```
[176]: array(['rt woman complain cleaning house man always take trash',  
        'rt boy dat cold tyga dwn bad cuffin dat hoe 1st place',  
        'rt dawg rt 0sbaby4life ever fuck bitch start cry confused shit',  
        ..., 'young buck wanna eat dat nigguh like aint fuckin dis',  
        'youu got wild bitch tellin lie',  
        'ruffled ntac eileen dahlia beautiful color combination pink orange  
yellow white coll'],  
        dtype=object)
```

```
[177]: labels
```

```
[177]: array([2, 1, 1, ..., 1, 1, 2], dtype=int64)
```

Tokenization and Padding

```
[178]: max_len = data['pre_process_tweet'].str.len().max()
```

```
[179]: max_len
```

```
[179]: 138
```

```
[180]: tokenizer = Tokenizer()  
tokenizer.fit_on_texts(tweets)
```

```
[181]: tokenizer.word_docs
```

```
[181]: defaultdict(int,  
        {'house': 130,  
         'rt': 7156,  
         'take': 317,  
         'man': 529,  
         'trash': 1115,  
         'woman': 247,  
         'always': 279,  
         'cleaning': 6,  
         'complain': 28,  
         'dwn': 2,  
         'tyga': 16,  
         'boy': 261,  
         'hoe': 4001,  
         '1st': 37,  
         'dat': 292,  
         'cuffin': 10,  
         'bad': 548,  
         'cold': 64,
```


'place': 75,
'Osbaby4life': 29,
'start': 158,
'fuck': 1363,
'confused': 27,
'shit': 1221,
'ever': 262,
'bitch': 10687,
'cry': 110,
'dawg': 43,
'tranny': 49,
'like': 2595,
'look': 562,
'faker': 3,
'ya': 432,
'hear': 64,
'told': 144,
'true': 65,
'might': 110,
'blow': 38,
'claim': 38,
'fucking': 607,
'faithful': 21,
'still': 466,
'somebody': 70,
'got': 1213,
'going': 267,
'much': 217,
'hate': 475,
'sit': 67,
'another': 169,
'girl': 665,
'cause': 297,
'skinny': 44,
'tired': 64,
'coming': 71,
'big': 231,
'u': 1195,
'get': 1470,
'back': 451,
'thats': 139,
'include': 5,
'mariam': 2,
'hobby': 9,
'fighting': 21,
'walked': 21,
'smh': 195,

'lol': 1051,
'everyone': 92,
'curve': 13,
'conversation': 15,
'keeks': 1,
'gang': 21,
'land': 14,
'murda': 5,
'yea': 85,
'smoke': 72,
'go': 674,
'ig': 40,
'loser': 16,
'thing': 252,
'nigga': 1793,
'miss': 77,
'whatever': 46,
'plz': 11,
'love': 688,
'everyday': 46,
'cut': 117,
'b': 156,
'black': 222,
'bottle': 26,
'cant': 99,
'nothing': 149,
'tell': 329,
'broke': 140,
'nino': 4,
'cancel': 9,
'wont': 24,
'change': 72,
'see': 461,
'even': 386,
'video': 79,
'dick': 342,
'bout': 322,
'kermit': 2,
'suck': 128,
'dont': 272,
'floor': 23,
'hardwood': 1,
'toeing': 3,
'tip': 25,
'door': 48,
'lip': 42,
'heaven': 13,

'pussy': 2140,
'hitting': 24,
'gave': 70,
'ocean': 8,
'met': 43,
'pill': 19,
'dr': 12,
'hennessy': 1,
'need': 518,
'trippy': 4,
'money': 272,
'want': 621,
'business': 46,
'spend': 24,
'txt': 4,
'old': 195,
'wetter': 4,
'new': 275,
'would': 366,
'say': 551,
'im': 428,
'excited': 19,
'shut': 174,
'murder': 18,
'game': 258,
'aint': 239,
'done': 127,
'toe': 21,
'stink': 14,
'there': 10,
'app': 10,
'bird': 567,
'bitter': 26,
'wrap': 11,
'angry': 28,
'son': 108,
'mane': 29,
'jus': 94,
'as': 1460,
'meet': 56,
'lame': 55,
'tear': 17,
'clown': 17,
'94': 3,
'snoop': 5,
'said': 363,
'cat': 66,

'inside': 26,
'doghouse': 1,
'momma': 44,
'hated': 9,
'favorite': 78,
'south': 19,
'yello': 1,
'body': 82,
'em': 178,
'nice': 89,
'naughty': 4,
'real': 382,
'chick': 99,
'make': 573,
'thick': 37,
'voice': 57,
'future': 29,
'pimp': 25,
'200': 10,
'picture': 92,
'post': 48,
'drug': 52,
'powerful': 2,
'call': 339,
'quick': 50,
'piece': 37,
'drive': 53,
'running': 46,
'round': 49,
'brand': 11,
'fucked': 167,
'send': 62,
'guy': 243,
'worst': 37,
'talk': 317,
'trust': 111,
'yall': 207,
'never': 428,
'cuffing': 10,
'bullet': 15,
'thank': 72,
'pregnant': 64,
'god': 151,
'dodge': 5,
'bound': 11,
'da': 263,
'chamberlain': 2,

'young': 103,
'music': 80,
'twitter': 251,
'cleaner': 2,
'teanna': 1,
'probably': 92,
'trump': 2,
'tonight': 129,
'gay': 138,
'act': 130,
'wait': 104,
'pitbull': 3,
'replace': 4,
'steal': 29,
'ima': 126,
'n': 254,
'gone': 177,
'movie': 65,
'tic': 1,
'took': 85,
'tacs': 1,
'wanna': 328,
'model': 25,
'russell': 5,
'simmons': 2,
'retweet': 23,
'mind': 81,
'deal': 31,
'teacher': 38,
'school': 169,
'fag': 283,
'stuff': 25,
'dating': 17,
'female': 147,
'cute': 122,
'think': 482,
'lad': 12,
'woof': 8,
'scally': 15,
'hot': 116,
'simplyaddictedtoguys': 1,
'yes': 130,
'asked': 47,
'kanye': 21,
'swear': 131,
'ur': 227,
'know': 827,

'side': 181,
'gf': 13,
'friend': 200,
'bring': 108,
'sol': 24,
'eat': 252,
'dish': 10,
'one': 722,
'oreo': 112,
'lemmie': 1,
'maybe': 77,
'better': 276,
'red': 108,
'please': 122,
'use': 124,
'could': 190,
'complete': 20,
'sentence': 19,
'niggah': 254,
'malone': 1,
'eggplant': 9,
'emoji': 14,
'looked': 30,
'lmao': 380,
'scream': 17,
'doe': 36,
'sexy': 54,
'venom': 1,
'gt': 9,
'hennessey': 3,
'shower': 19,
'curtain': 2,
'seh': 1,
'racist': 64,
'pic': 126,
'posted': 16,
'every': 282,
'skateboarding': 1,
'faggot': 516,
'pushing': 6,
'ball': 79,
'yo': 568,
'wood': 13,
'life': 316,
'monkey': 158,
'awwwwe': 1,
'baby': 176,

'adorable': 9,
'bathtime': 1,
'soooo': 12,
'skippy': 3,
'damn': 319,
'sidewalk': 2,
'chance': 29,
'kill': 135,
'must': 78,
'ant': 4,
'way': 228,
'30': 43,
'mvp': 6,
'0': 41,
'ring': 28,
'hood': 49,
'kid': 154,
'funnnnnnn': 1,
'ppl': 59,
'growing': 7,
'ghetto': 262,
'fact': 73,
'tattoo': 22,
'nipple': 19,
'tattooing': 1,
'wanting': 13,
'cuteeeee': 1,
'worshiping': 2,
'lost': 75,
'buttcheek': 1,
'piercings': 4,
'wow': 72,
'yeah': 217,
'shrug': 5,
'ok': 99,
'si': 10,
'work': 186,
'rather': 65,
'queer': 93,
'gaywad': 1,
'funny': 121,
'tweet': 201,
'read': 54,
'k': 44,
'ungrateful': 9,
'fr': 17,
'lulwhore': 1,

'kidnapped': 2,
'beaner': 33,
'mexican': 32,
'he': 20,
'alsarabsss': 1,
'happy': 166,
'nigs': 26,
'birthday': 80,
'thanks': 88,
'homework': 11,
'eating': 98,
'paul': 13,
'john': 25,
'hell': 146,
'stop': 327,
'weave': 22,
'lil': 258,
'bald': 15,
'ugly': 240,
'headed': 19,
'making': 112,
'leaf': 18,
'riot': 3,
'looting': 1,
'win': 86,
'nig': 112,
'attend': 6,
'9': 29,
'year': 211,
'car': 124,
'retard': 139,
'come': 363,
'food': 72,
'prople': 1,
'show': 157,
'richnow': 1,
'brought': 27,
'tinder': 3,
'hella': 52,
'8': 52,
'anymore': 45,
'chill': 69,
'cowboy': 15,
'shutup': 7,
'lmaoooooooooooo': 2,
'depressing': 3,
'foh': 22,

'lbvs': 2,
'g': 41,
'spitting': 6,
'trick': 31,
'jail': 33,
'cuzzo': 2,
'couple': 61,
'page': 21,
'let': 375,
'excuse': 35,
'harlem': 6,
'second': 51,
'shake': 20,
'full': 88,
'miley': 17,
'damaged': 1,
'amen': 5,
'bc': 84,
'emotionally': 1,
'wayyyyy': 2,
'idc': 26,
'loyal': 133,
'meow': 2,
'licking': 3,
'check': 58,
'stand': 78,
'12th': 1,
'fan': 101,
'ny': 12,
'threw': 27,
'foothill': 1,
'happen': 26,
'vixen': 1,
'ran': 31,
'ent': 3,
'okay': 75,
'pat': 11,
'nude': 48,
'holding': 11,
'anyway': 34,
'blacklisted': 1,
'nobody': 108,
'wrong': 125,
'type': 93,
'top': 94,
'3': 162,
'dm': 40,

'baseball': 20,
'season': 52,
'started': 49,
'stupid': 231,
'little': 314,
'yoouoooooooo': 1,
'dead': 79,
'serious': 53,
'wheet': 1,
'tho': 245,
'really': 415,
'wise': 10,
'worm': 24,
'early': 61,
'owl': 2,
'night': 177,
'sound': 97,
'overdosing': 1,
'day': 431,
'heavy': 5,
'creature': 6,
'fairy': 23,
'photographed': 1,
'lovely': 7,
'tale': 8,
'pit': 12,
'bull': 17,
'aw': 16,
'lakers': 15,
'long': 116,
'careless': 2,
'cc': 8,
'hahaha': 47,
'kiss': 38,
'anything': 105,
'soccer': 17,
'yup': 10,
'haircut': 9,
'khaki': 6,
'cargo': 6,
'coloured': 12,
'cheat': 57,
'wear': 94,
'play': 197,
'lap': 7,
'fat': 208,
'17': 16,

'screamn': 1,
'finna': 61,
'missed': 20,
'p': 28,
'thinkin': 20,
'ask': 124,
'rule': 18,
'cousin': 26,
'gotta': 307,
'lmmfao': 2,
'anybody': 20,
'moving': 17,
'jersey': 14,
'amo': 3,
'los': 7,
'pie': 12,
'ainna': 1,
'meat': 14,
'slap': 52,
'thinking': 58,
'bruh': 136,
'wit': 333,
'edge': 7,
'talking': 241,
'tooth': 1,
'treating': 7,
'bag': 59,
'sweet': 44,
'nah': 124,
'bandit': 1,
'came': 68,
'mom': 131,
'hacked': 4,
'right': 381,
'sex': 124,
'open': 61,
'lmfao': 95,
'watch': 130,
'eye': 92,
'shoulda': 17,
'coulda': 11,
'woulda': 10,
'roll': 38,
'preciate': 1,
'bday': 18,
'used': 78,
'el': 12,

'duh': 12,
'wcw': 9,
'pass': 4,
'hacky': 1,
'throw': 102,
'sack': 11,
'fuckin': 250,
'atlanta': 8,
'dumb': 257,
'getting': 235,
'bunch': 49,
'mad': 199,
'outta': 77,
'glad': 43,
'smart': 24,
'wife': 79,
'cheating': 31,
'first': 177,
'force': 12,
'cuz': 190,
'relationship': 67,
'care': 142,
'single': 89,
'power': 30,
'white': 364,
'blunts': 7,
'bro': 136,
'spit': 15,
'dude': 178,
'spark': 4,
'blowed': 1,
'drunk': 65,
'around': 192,
'world': 112,
'sucked': 11,
'party': 85,
'set': 46,
'lot': 119,
'cartoon': 6,
'captain': 13,
'saving': 16,
'alway': 1,
'saveahoe': 1,
'nose': 23,
'leg': 44,
'sox': 24,
'bois': 2,

'brother': 53,
'wish': 117,
'keezy': 1,
'ooooow': 2,
'bet': 95,
'dressed': 16,
'halloween': 37,
'basic': 83,
'pornhub': 2,
'gaming': 2,
'stream': 4,
'feature': 6,
'gtfo': 13,
'katie': 1,
'choke': 23,
'scared': 35,
'hit': 237,
'bed': 59,
'ha': 49,
'well': 201,
'iguess': 1,
'theyll': 1,
'oh': 190,
'molly': 18,
'5': 104,
'ho': 192,
'seed': 5,
'refer': 12,
'beating': 16,
'finally': 32,
'peel': 2,
'move': 72,
'tf': 108,
'shaq': 3,
'line': 81,
'rewind': 1,
'alu': 4,
'hahahaha': 22,
'record': 18,
'sai': 1,
'pe': 3,
'haha': 149,
'vine': 20,
'anywaaaaaays': 1,
'io': 10,
'mixtape': 16,
'waiting': 40,

'mayne': 6,
'cover': 15,
'naw': 65,
'tape': 15,
'upcoming': 1,
'dirty': 65,
'foot': 44,
'ol': 39,
'amazing': 18,
'yellow': 253,
'flag': 13,
'niggy': 1,
'time': 483,
'iffy': 5,
'actin': 31,
'special': 34,
'nation': 19,
'alive': 20,
'stay': 134,
'knight': 5,
'smfh': 16,
'decent': 8,
'wonder': 62,
'many': 171,
'stfu': 101,
'screamed': 5,
'bye': 47,
'keep': 238,
'soft': 24,
'stud': 9,
'blame': 25,
'dry': 27,
'condom': 10,
'dropping': 8,
'gucci': 24,
'text': 113,
'month': 55,
'mixtapes': 2,
'wats': 5,
'burbs': 2,
'issue': 21,
'sounding': 6,
'daddy': 37,
'mewtwo': 3,
'holographic': 1,
'finding': 10,
'without': 76,

'cum': 22,
'easy': 36,
'sheen': 32,
'major': 16,
'respect': 59,
'charlie': 282,
'bae': 72,
'twerking': 13,
'cook': 25,
'wtf': 113,
'good': 506,
'along': 14,
'zoe': 2,
'fight': 120,
'snake': 10,
'everywhere': 22,
'booty': 66,
'sheesh': 3,
'update': 12,
'report': 16,
'tuned': 2,
'1b': 2,
'week': 97,
'program': 8,
'rft': 2,
'oregon': 3,
'field': 25,
'terrell': 1,
'best': 187,
'tv': 32,
'romo': 8,
'cried': 8,
'eagle': 12,
'owen': 1,
'def': 17,
'people': 399,
'lmfaoooo': 15,
'nigger': 310,
'potent': 2,
'lick': 37,
'slept': 18,
'past': 42,
'pissed': 19,
'w': 150,
'frat': 4,
'arch': 1,
'feel': 232,

'fam': 36,
'depends': 4,
'hick': 25,
'confusing': 4,
'america': 43,
'santi': 1,
'mean': 218,
'nasty': 69,
'toto': 1,
'vagina': 17,
'choose': 10,
'actually': 76,
'vibe': 13,
'dis': 115,
'tried': 69,
'grim': 1,
'tight': 30,
'shirt': 46,
'everybody': 110,
'af': 85,
'saying': 108,
'may': 50,
'grandkids': 1,
'10': 82,
'climate': 3,
'beautiful': 46,
'gurls': 1,
'gurl': 5,
'purple': 15,
'r': 83,
'sum': 47,
'playin': 35,
'track': 20,
'ceeeleee': 1,
'shitted': 2,
'taller': 7,
'blurry': 1,
'mine': 59,
'seen': 137,
'obsessed': 7,
'meant': 18,
'kobe': 20,
'forget': 47,
'least': 60,
'asking': 35,
'asian': 50,
'mixed': 9,

'lmaoo': 45,
'hero': 8,
'roach': 7,
'cripple': 25,
'ariza': 1,
'coward': 15,
'foster': 3,
'imaginary': 5,
'whoa': 8,
'home': 130,
'everything': 89,
'lowkey': 20,
'sleep': 92,
'quiet': 13,
'weird': 41,
'bringing': 11,
'mention': 32,
'weak': 49,
'homeboy': 11,
'advice': 14,
'speed': 5,
'animal': 29,
'quote': 27,
'stripper': 17,
'dancer': 1,
'pop': 70,
'ballet': 2,
'beethoven': 2,
'beer': 39,
'andrewbryant': 1,
'vision': 5,
'poff': 1,
'2': 254,
'double': 30,
'computer': 8,
'twerk': 14,
'prince': 2,
'takin': 14,
'art': 9,
'welfare': 7,
'aid': 18,
'government': 11,
'claimed': 1,
'slander': 3,
'61': 2,
'role': 10,
'tree': 24,

'find': 150,
'lime': 2,
'hiding': 10,
'glitter': 1,
'shoe': 35,
'nike': 15,
'jordan': 17,
'drake': 34,
'released': 10,
'di': 5,
'inna': 1,
'alone': 46,
'austin': 17,
'teach': 16,
'curved': 8,
'chose': 9,
'expects': 2,
'retweetin': 1,
'naked': 33,
'lmaooo': 39,
'cyrus': 12,
'mavs': 1,
'tony': 19,
'cowgirl': 2,
'fuk': 23,
'chocolate': 28,
'highmore': 1,
'holy': 26,
'laughing': 33,
'factory': 4,
'freddie': 2,
'using': 45,
'anonymous': 5,
'whata': 1,
'follower': 41,
'annoying': 51,
'clearly': 25,
'bitching': 11,
'catfish': 9,
'catching': 13,
'slick': 12,
'worried': 41,
'worry': 28,
'insecure': 13,
'handle': 30,
'tent': 3,
'scooting': 1,

'caring': 7,
'sharing': 13,
'share': 17,
'problem': 96,
'ionn': 2,
'hungry': 40,
'dese': 23,
'lyin': 7,
'eli': 5,
'interception': 1,
'manning': 8,
'27th': 2,
'coach': 18,
'senior': 16,
'brown': 70,
'sarkisian': 1,
'anthony': 6,
'head': 178,
'usc': 2,
'team': 97,
'steve': 14,
'rb': 2,
'quits': 1,
'mansplaining': 1,
'zero': 19,
'adios': 1,
'also': 68,
'guess': 98,
'chilling': 9,
'impress': 14,
'simple': 21,
'rigghhht': 1,
'control': 35,
'idk': 68,
'tuku': 3,
'blah': 12,
'cuh': 27,
'ita': 2,
'suure': 1,
'crazy': 159,
'slice': 4,
'gonna': 281,
'kevin': 10,
'hart': 6,
'fine': 63,
'dam': 23,
'oomf': 21,

'magaluf': 1,
'sends': 6,
'drink': 56,
'proably': 1,
'24': 33,
'later': 46,
'c': 40,
'pinkman': 1,
'jesse': 2,
'hickies': 1,
'thigh': 8,
'longer': 21,
'front': 54,
'friday': 37,
'ipad': 2,
'yesterday': 28,
'supahot': 1,
'drop': 61,
'test': 28,
'fohead': 1,
'chrus': 2,
'walk': 79,
'forever': 32,
'med': 3,
'slush': 2,
'taste': 54,
'bobby': 17,
'live': 105,
'workout': 11,
'stoner': 8,
'bar': 38,
'ride': 54,
'traphoussseeeee': 1,
'sad': 55,
'15': 29,
'character': 13,
'stark': 1,
'death': 34,
'harbor': 3,
'bus': 22,
'promo': 2,
'ultraviolence': 1,
'bitchy': 5,
'lana': 5,
'del': 2,
'rey': 5,
'kittycopter': 1,

```

'highway': 2,
'smuggler': 1,
'explains': 5,
'fighter': 12,
'jihadi': 63,
'helped': 7,
'avis': 2,
'aww': 6,
'together': 61,
'avi': 32,
'ike': 13,
'exposed': 10,
'ont': 5,
'exactly': 24,
'boosting': 2,
'healthy': 4,
'why': 2,
'size': 16,
'kim': 30,
'protect': 14,
'stalker': 3,
'loco': 4,
'dope': 33,
'near': 23,
'pullin': 4,
'taught': 17,
'aye': 40,
'girlfriend': 79,
'marack': 1,
'bruno': 4,
'beefing': 8,
'sike': 6,
'bottom': 17,
'9th': 5,
'rally': 8,
'advantage': 3,
'everyjuan': 1,
'person': 76,
'pal': 6,
'till': 51,
'knowing': 25,
'stack': 8,
...})

```

```

[182]: sequences = tokenizer.texts_to_sequences(tweets)
sequences

```

[182]: [[2, 93, 886, 2969, 196, 28, 76, 61, 13],
 [2, 88, 67, 424, 1413, 5837, 27, 2069, 67, 3, 695, 370],
 [2, 618, 2, 860, 90, 10, 1, 169, 239, 925, 11],
 [2, 25, 4, 550],
 [2, 11, 419, 228, 412, 228, 4639, 1, 177, 36],
 [11, 681, 682, 1144, 394, 35, 21, 3],
 [404, 34, 156, 1, 12, 117, 11, 83],
 [69, 425, 107, 1, 395, 9, 573, 19],
 [228, 8, 36, 1, 38, 185],
 [2243, 3333, 1145, 5838, 1],
 [8182, 1, 1571, 302, 14, 1109, 1481, 4, 128],
 [3334, 1035, 1, 1035, 1572],
 [3, 390, 1414, 325, 18, 656],
 [27, 1, 99, 4],
 [1, 8],
 [1, 6, 339],
 [1, 1782, 574],
 [1, 17],
 [1, 8, 234, 575, 165],
 [114, 956, 27, 1],
 [180, 1, 279, 57, 174],
 [2244, 1, 4, 3335],
 [279, 37, 3, 986, 380],
 [10, 1, 70, 46, 200, 53, 4640, 349, 58, 10, 656],
 [12, 36, 1, 987, 4641, 8183, 1079],
 [5, 619, 4, 1671, 529],
 [3, 1036],
 [620, 5, 2245, 1783, 396, 5, 1231],
 [30, 3870, 1, 10, 8184],
 [1037, 79, 20, 1, 555],
 [3871, 123, 1, 72, 1, 5, 3872],
 [26, 39, 38, 123, 123, 1, 51, 8, 1232],
 [98, 58, 1289, 86, 5, 6, 157],
 [988, 209, 5, 1573],
 [39, 209, 926, 1, 1933, 887, 24, 2070, 2071],
 [36, 1],
 [98, 174, 234, 1],
 [280, 477, 253, 36, 861, 7, 1],
 [442, 239, 3, 185, 1358, 1359],
 [4, 3336, 50, 4642, 70, 17, 3],
 [600, 50, 5, 406, 927, 8185],
 [2246, 3, 356],
 [311,
 19,
 27,
 23,
 8,

3873,
 27,
 8186,
 3,
 44,
 311,
 334,
 1186,
 281,
 4,
 130,
 44,
 683],
 [957, 957, 3, 862, 467],
 [530, 290, 5, 8, 2072, 4],
 [5, 5839, 519],
 [531, 696, 5, 54, 499],
 [576, 542, 4, 1934, 72, 5, 185, 58, 8, 158],
 [1, 46, 697, 443, 102],
 [3, 4, 6, 1037, 79, 60, 58],
 [70, 225, 6, 1],
 [112, 6, 165, 2073, 3, 69, 112, 98, 41, 1],
 [3337, 1482, 66, 1, 234, 420, 1935, 384, 170],
 [269, 1231, 5840, 1, 17, 335],
 [8187, 5841, 303, 5842, 97, 3],
 [183, 6, 254, 202, 4, 1, 205],
 [210, 839, 406, 78, 3874, 4643, 2, 860, 1],
 [137, 76, 3],
 [253, 1, 326, 8188, 8189, 4, 421],
 [56, 8, 79, 10, 958, 1, 2694, 5843],
 [303, 336, 149, 117, 815, 1, 7, 665, 1038],
 [21, 77, 50],
 [175, 32, 1290, 5, 216, 989, 23, 5],
 [8190, 1415, 1415, 236, 1416, 1483],
 [36, 90, 556, 1, 1, 1080, 467, 197],
 [9, 137, 255, 103, 141, 1, 86, 15, 36, 1672, 119, 206],
 [1415, 1415, 236, 990],
 [8191, 85, 240, 2074, 16, 240, 14],
 [1, 10, 9, 337, 8, 81],
 [10, 248, 8192, 28, 1, 7, 82, 135, 222, 214, 1187, 1233],
 [2247, 1574, 717, 26, 816, 4, 1360, 47],
 [4644, 8193, 2248, 16, 478, 1],
 [8194, 5, 3872, 1188, 5844, 47],
 [74, 30, 37, 203, 1417, 1, 50, 405, 11],
 [22, 10, 8195, 1673, 2970, 31, 28, 343, 62],
 [143, 160, 8196, 8197, 1784, 2249],
 [11, 63, 4645, 14],

[9, 90, 190, 3875, 5845, 32, 63, 6, 110, 8, 5, 165, 362, 165, 601, 840],
 [602, 888, 602, 2971, 602, 1, 14],
 [461, 60, 27, 89, 551, 163, 2695, 6, 8198],
 [371, 1081, 1081, 5, 8199, 1234],
 [2, 1, 1674, 202, 4, 6, 27, 11, 98, 8200, 6, 202, 4, 1],
 [8, 5846, 1, 1415, 1415],
 [381, 2, 372, 8201, 3876, 116, 3],
 [275, 1936, 407, 180, 1, 3338, 57, 138],
 [300, 8202],
 [31, 490, 122, 67, 603, 46, 229, 14],
 [1, 2250, 1291, 8203],
 [5847, 22, 1],
 [8204, 1110, 741, 128, 57, 1110, 758],
 [148, 340, 959, 312, 22],
 [103, 665, 57, 9, 9, 1937, 1, 30, 18],
 [6, 276, 3, 14, 179, 325, 14, 889, 1575, 6, 50, 889, 1575],
 [51, 84, 100, 1418, 1235, 1, 59, 249, 1082],
 [1292, 81, 320, 63, 86, 18, 4646, 11, 81, 169, 8205, 242],
 [83, 38, 149, 200, 53, 3, 2972],
 [110, 10, 22, 1, 863, 115, 123],
 [52, 41, 255, 363, 70, 213, 184],
 [8206, 167, 520, 4647, 3, 39, 119, 588, 398, 928, 4, 4, 507, 8207],
 [58, 1419, 2696, 5],
 [5848, 6, 1, 20, 4, 4648, 11, 1111],
 [19, 15, 2973, 646, 11, 83, 5849, 1, 1782],
 [48, 23, 408, 122, 5850, 1146, 8, 97, 718, 47, 1, 796],
 [2974, 1147, 742, 18, 291, 184, 601, 532],
 [3339, 1361, 30, 202, 4, 3, 328, 8208, 8209, 1111],
 [16, 198, 175, 5851, 81, 3, 929],
 [5, 4649, 5, 4650, 4650],
 [453, 8210, 28, 46, 31, 1785, 265, 350],
 [960, 8211, 4651, 1, 12, 817, 890, 141, 4, 8212, 1],
 [39, 1, 368, 500, 1938, 381],
 [21, 183, 8213, 3, 1939, 743],
 [10, 217, 1, 252, 15],
 [112, 157, 23, 1],
 [34, 420, 292, 1],
 [12, 1, 621, 90, 301, 155],
 [1189, 501, 320, 17, 552],
 [64, 109, 1, 10, 8214],
 [42, 556, 357, 7, 502, 104, 135, 26, 1, 8215, 14],
 [444, 24, 142, 5852, 2075, 991],
 [8216, 3340, 519, 285, 27, 205, 5, 11, 74, 40],
 [1786, 1293, 8217, 2697, 1039, 2251, 2698, 1039, 2251, 2698],
 [1362, 77, 122, 1362, 14],
 [241, 1484, 13, 135, 5853, 44, 2453, 563],
 [26, 266, 104, 666, 31],

[126, 1363, 445, 19, 282, 2975, 1787, 2976, 2252, 1940],
 [135, 18, 118, 7, 1, 2699],
 [1420, 1190, 54, 66, 22, 841, 1, 97, 446, 8218],
 [36, 141, 1, 68, 15, 1294, 11, 1191, 218, 891, 1364, 1148, 58, 1365, 5854],
 [155, 4652, 2700, 1576, 10, 116],
 [3, 471, 1485, 521, 7, 8219, 186],
 [112, 1, 55, 2701, 101],
 [112,
 35,
 83,
 796,
 2702,
 115,
 211,
 179,
 116,
 210,
 66,
 589,
 8220,
 8221,
 839,
 1,
 437,
 78,
 11],
 [56, 223, 195, 10, 45, 5, 399, 63, 12, 3877, 4, 10],
 [666, 9, 296, 454, 201, 1, 297],
 [1941, 1942, 1366, 7, 3],
 [148, 1192, 1, 7, 6, 8222, 1, 684],
 [344, 443, 1],
 [100, 1, 4, 2253, 1788, 1, 252, 1577],
 [6, 15, 8223, 1943, 3, 211, 270, 25, 3878, 47],
 [622, 108, 364, 2454, 174, 553, 6, 364, 718, 92, 1, 125, 163, 95, 130],
 [89, 992, 929],
 [51, 351, 3, 133, 182, 797, 1789, 409, 150, 370],
 [304, 742, 3],
 [49, 88, 798, 1, 114, 798, 1],
 [48, 8, 422, 8, 8224, 1486, 11, 3879, 2703, 10, 3, 191, 49, 144, 398],
 [1, 9, 131, 250, 38],
 [8, 53, 1944, 22, 305, 27, 1, 1291, 137, 557],
 [1784, 216, 303, 3, 226],
 [74, 2977, 8225, 1367, 3, 47, 47, 63, 1675, 8226],
 [20, 8, 1040, 27, 5855, 236, 993, 604],
 [215, 1, 4, 491, 8227, 5856, 286],
 [1421, 406, 698, 313, 5],
 [5857, 8, 5858, 3880, 2704, 92, 1, 1676, 8, 8228],

[59, 719, 1083, 223, 9, 1, 7, 6, 4, 26, 102, 105, 93, 447, 1111],
 [455, 124, 8229, 23, 3, 198, 191, 8230, 41, 198, 3, 242],
 [132, 28, 251, 1236, 121, 10, 36, 3341],
 [28, 1790, 74, 19, 1],
 [297, 406, 552, 1422, 11, 6, 799, 12, 5, 6],
 [4653, 4653, 255, 38, 8231, 38, 385, 4654, 333, 5856, 258, 50, 13],
 [1945, 1945, 3342, 1237, 121, 1149, 8232, 4655, 1112, 14, 8233, 176],
 [426, 123, 1487, 72, 1423, 8234, 1424, 116, 657, 3, 2978],
 [667, 413, 577, 7, 1, 229],
 [8, 94, 1578, 1295, 179, 116, 8235],
 [12, 33, 1, 800, 3343],
 [52, 187, 1193, 3344, 1238, 35, 444, 24, 759],
 [3, 198, 1425, 438, 252, 2455, 20],
 [186,
 2254,
 106,
 471,
 1041,
 179,
 218,
 36,
 19,
 153,
 33,
 3345,
 1,
 132,
 368,
 272,
 533,
 3],
 [892, 5, 1, 76, 893, 2076],
 [157, 22, 1, 7],
 [994, 861, 718, 2456, 5859, 74, 492, 3, 46, 235, 38],
 [17, 3, 79, 3346, 217, 6],
 [198, 1, 5860, 358, 685, 1150, 4, 2077, 8236, 4656, 2705, 4, 14],
 [699, 1113, 699, 3, 371],
 [75, 760, 44, 1426, 21, 431, 75, 760],
 [215, 359, 12, 3, 104],
 [10, 1677, 1, 995],
 [1, 245],
 [157, 4657, 219, 5, 4657, 29, 28, 29, 385, 1579],
 [47, 21, 2078, 1, 130],
 [107, 386, 3, 1114, 4658, 63],
 [187, 5861, 1791, 1427, 5862, 287],
 [126, 1363, 445, 19, 282, 2975, 1787, 2976, 2252, 197, 297],
 [3881, 1083, 74, 115, 8, 996, 13, 7, 2457, 5863],

[1368,
 1419,
 14,
 2,
 8237,
 8238,
 139,
 1792,
 1419,
 90,
 144,
 2458,
 4,
 1,
 761,
 2459],
 [1488, 34, 114, 43, 114, 43, 65],
 [135, 9, 1, 5864],
 [1194, 1483, 633, 51, 686, 413, 990, 1296],
 [17, 3882, 88, 171, 1041, 241, 990],
 [478, 1483, 236, 990, 8239],
 [211, 700, 68, 445, 3, 3883, 175, 111],
 [12, 997, 14],
 [72, 97, 3884, 11, 18, 38, 1186, 605, 1],
 [5865, 113, 7, 1369, 32, 5, 69, 5865, 8240, 4, 391, 5],
 [373, 1946, 100, 1, 7, 362, 118],
 [1, 190, 1580, 36, 400, 230, 17, 1361],
 [425, 1, 256, 25, 113, 6, 107, 330, 282, 842, 564, 8241, 231],
 [331, 24, 8242, 493, 41, 37, 312, 4659, 380],
 [187, 565, 1],
 [321, 1489, 8243, 28, 3347, 280, 720, 109, 3, 1151, 35, 5866, 543, 8244],
 [79, 108, 2706, 1, 108, 8245],
 [41, 192, 1, 2255, 6, 2255, 432],
 [1152, 187, 101, 13, 566, 1297, 16, 128],
 [462, 65, 297],
 [43, 76, 721, 2256, 27, 544, 1, 14],
 [8246, 2078, 1490, 462, 961, 4, 2460, 2707, 578],
 [4660, 207, 3348, 13, 2257, 398, 167, 306, 258],
 [1115, 425, 425, 1678, 18, 293, 3],
 [59, 1947, 623, 65, 801],
 [17, 3],
 [22, 19, 279, 1948, 667, 534, 1, 7, 6],
 [32, 697, 1581, 396, 5, 3349, 14],
 [864, 122, 356, 843, 5],
 [34, 1298, 57, 5867, 3, 4661, 5867, 382, 5, 5868, 1298],
 [8247, 5869, 818, 2979, 87, 634, 5869, 818, 5, 128],
 [1, 1370, 2461, 36, 36, 63],

[5870, 2258, 252, 1582, 1],
 [5871, 2708, 1949, 1299, 8248, 49, 43, 114, 4662, 13],
 [1, 182, 126, 2079],
 [326, 173, 2080, 5872, 998, 230, 1, 66, 217, 998, 191],
 [722, 72, 647, 2081, 1491, 1300, 197, 8249, 647, 144, 77],
 [8250, 3350, 89, 567, 578],
 [3, 1301, 1371, 9, 1301, 1371, 9, 8, 2462, 15],
 [1, 995, 15, 200, 53, 16],
 [118, 3, 8, 2259, 14],
 [124, 3, 16, 5873, 198],
 [667, 1, 7],
 [186, 39, 425, 6, 8251, 1361, 1793, 778, 1, 98, 292, 27, 668],
 [1239, 2459, 8252, 86, 70, 113, 35, 1, 47, 1042, 5874, 14, 1, 10],
 [962, 11, 5875, 8253, 75, 844, 3885, 684, 1079, 779],
 [20, 223, 1672, 111, 4, 590, 3351, 8254, 31],
 [16, 624, 394, 1],
 [36, 1, 76, 1950, 999, 182, 535, 1],
 [2260, 3, 1148, 1794, 6, 1679],
 [865, 1, 591, 58],
 [141, 3, 1680, 6, 802, 44, 16],
 [2, 47, 1, 52, 8255, 4663, 4, 10, 11],
 [891, 50, 1372, 3, 491, 50, 1681, 2709, 17, 491],
 [648, 5876, 56, 1037, 79, 363, 180, 3, 273],
 [1043, 3, 2710, 9, 6],
 [132, 116, 10, 1],
 [3352, 2463, 890, 5877, 8256, 501, 47, 13],
 [5878, 1373, 3886, 2980, 392, 54, 147, 1240, 1583, 8257, 405, 8258, 277, 1],
 [288, 410, 1241, 33, 8259, 8260, 272, 5],
 [2261, 478, 1416, 1483],
 [1681, 1, 432, 432],
 [79, 1584, 1153, 180, 1],
 [41, 1, 48, 43, 744, 380, 8261],
 [4664, 121, 393, 930, 930, 930],
 [125, 931, 563, 8262, 4664, 121, 5879],
 [3, 164],
 [90, 73, 3887, 1, 2082, 2981],
 [1154, 95, 427, 104, 1084, 1],
 [438, 19, 396, 147, 762, 88, 472, 8263, 97, 8264, 2982, 500, 77],
 [116, 1, 116, 1, 54, 337, 165, 658, 579, 5880, 8265],
 [8266, 2464, 45, 5, 45],
 [51, 35, 17, 1155, 269, 565, 100, 160, 7],
 [52, 479, 1, 395],
 [31, 1],
 [325, 5881, 138, 687, 343, 31, 138],
 [1416, 1483, 51, 8],
 [698, 894, 109, 65],
 [5882, 35, 8267, 104, 360, 665, 819, 448, 819, 1, 8268],

[17, 276, 5, 14, 41, 8, 291, 14, 85, 11, 763, 14, 90],
 [5883, 3, 391, 743, 700, 3, 485, 4, 4665],
 [5883, 3, 391, 743, 700],
 [1242, 108, 263, 10, 179, 325, 126, 3, 1951],
 [22, 1, 994, 861, 2, 1, 2465],
 [503, 1, 802, 669],
 [3, 8, 8269],
 [866, 486, 761, 1682, 764, 35, 926, 8270, 28],
 [4666, 1, 305, 963, 3, 14],
 [5, 17],
 [8271, 349, 395, 64, 1, 3353, 5884, 3354, 2983, 5885],
 [110, 10, 1, 297, 8272],
 [8273, 3355, 2711, 1795, 1579, 433, 5886],
 [2984, 4667, 25, 216, 449, 68, 8, 1683, 765, 1],
 [1153, 70, 3, 3356, 865, 2083, 219],
 [5887, 1, 5888, 1374, 1044, 3888, 104],
 [286, 1, 845, 533],
 [63, 1045, 68, 1585, 122, 4668, 535, 12, 766, 53, 499, 1, 3889],
 [249, 8274, 352, 659, 165, 3890, 3, 4, 1302],
 [3891, 77, 88, 17],
 [362, 1, 2985, 426, 76, 2466],
 [25, 2467, 1303, 3357],
 [61, 4669, 311, 365, 3, 8275],
 [148, 340, 1, 6, 312, 2986],
 [486, 1000, 3, 10, 328, 12, 2262, 254, 536, 687],
 [4670,
 11,
 1952,
 670,
 6,
 4671,
 670,
 55,
 456,
 820,
 133,
 67,
 36,
 1948,
 670,
 161],
 [2263, 50, 134, 1243, 134, 1243, 490, 959],
 [1046, 44, 6, 136, 69, 5, 6, 73, 1046, 45, 38, 723, 8276, 1147],
 [5, 35, 1047, 145, 1001, 19, 26],
 [7, 2468, 16, 2264, 99, 176, 1586, 19, 7, 12, 125, 98, 8, 5, 287],
 [116, 2469, 117, 81, 21, 184, 37, 125],
 [457, 245, 964, 1, 1796, 222, 2987, 780, 3, 178, 4, 114, 5889],

[3358, 1, 30, 187, 322, 78, 8, 45, 10, 364, 136],
 [34, 88, 2712, 1, 128],
 [36, 959, 35, 256, 85, 745, 243, 136],
 [39, 746, 7, 1, 30, 404, 1492],
 [3],
 [1575, 1148, 58, 414, 5890, 66, 508, 8277, 11, 1, 1587, 187, 66, 468],
 [1, 509, 6, 8, 1428, 6, 18, 1002, 2265, 537, 580],
 [119, 34, 469, 3],
 [17, 10, 6, 3],
 [236, 1485, 478, 10, 1416, 1483, 51, 17, 477, 2988],
 [2470, 8278, 1416, 1483, 10, 116],
 [1, 17, 2084],
 [2713, 1244, 601, 1085, 532, 245, 3892, 69, 96, 128],
 [151, 406, 895, 747, 5],
 [1415, 1415, 478, 781, 88, 1416],
 [9, 37, 63, 188, 4672, 1048, 1, 23, 2471, 146],
 [8279, 356, 4673, 32, 229, 1110, 42, 578, 1110, 64, 3],
 [331, 223, 1953, 9, 54, 21, 724, 364, 1],
 [1588, 896, 216, 31, 473],
 [16, 28, 13, 156, 28, 2266],
 [12, 292, 27, 1, 99, 4, 8280],
 [8281, 2453, 1240, 20, 111, 8282, 8, 242, 1429],
 [170, 821, 867, 1797, 49, 1, 3359],
 [3, 3, 8283],
 [658, 54, 8284, 5891, 54, 1684, 1003, 605, 1156, 3],
 [73, 339, 1, 963, 1798, 332],
 [34, 1, 168, 91, 4674, 1, 1304, 1157, 91, 8285, 724],
 [4675, 269, 400, 2712, 24, 3360, 329, 563],
 [112, 8286, 1245, 1, 635, 14],
 [61, 503, 1415, 1415],
 [132, 170, 17, 4676, 8287, 1, 17],
 [241, 2989, 11, 803, 606, 1, 797, 868, 132],
 [45, 1, 128],
 [2472, 67, 5, 62, 68, 341, 8288, 688, 427],
 [897, 248, 3893, 607, 65, 114, 15, 412],
 [1116, 932, 5892, 932, 454, 24, 168, 660, 3894],
 [6, 181, 6, 6, 285, 95, 96],
 [349, 1117, 689, 2085, 283, 2990],
 [46, 822, 272, 8289, 195, 823, 5893, 5893, 804, 1589, 220, 2714],
 [2715, 5894, 12, 234, 74, 277, 1195, 224, 12, 3, 8290, 128],
 [581, 1, 2716, 1118],
 [95, 236, 933, 5, 10, 116],
 [433, 3895, 8291, 2086, 8292, 8293],
 [54, 8294, 69, 189, 3],
 [135, 41, 4677, 84, 53, 2991, 12, 2992, 134, 1, 132, 170],
 [201, 93, 168, 898, 4, 53, 373, 158, 5, 34, 1],
 [390, 314, 636, 10, 538, 28, 56, 307, 52, 96],

[8295, 24, 104],
 [5, 589, 4, 2473, 8296],
 [97, 824, 621, 13, 1799],
 [934, 1, 3896],
 [4678, 74, 1, 189, 3, 5895],
 [206, 1154, 168, 101, 522, 5],
 [1, 2474, 258, 164, 41, 1954],
 [603, 533, 3, 1800, 533],
 [8297, 184, 462, 23, 5896, 767],
 [210, 106, 487, 5, 7, 6, 480, 377, 34],
 [725, 568, 1671],
 [8298, 4679, 726, 1, 8299],
 [141, 1, 1243],
 [371, 8300, 1, 129, 11, 47],
 [10, 55, 9, 2475, 3, 3897],
 [800, 4, 2476, 40, 1305],
 [374, 84, 1, 690, 46, 111, 246, 1590],
 [21, 236, 17, 4680, 8301],
 [92, 7, 6, 1685, 3, 534, 10, 10, 314, 3, 240],
 [189, 22, 1, 8302],
 [225,
 1043,
 3,
 2477,
 177,
 8303,
 187,
 265,
 1306,
 434,
 3361,
 446,
 52,
 5891,
 463,
 1955],
 [659, 112, 210, 5897, 3898, 277, 2087, 5897, 1591, 104, 2070, 166, 184],
 [9, 12, 9, 20, 1, 1004, 164, 305, 51, 17, 2972],
 [159, 1246, 296, 8304, 432, 69, 37, 47],
 [371, 144, 3, 26, 12, 3],
 [5898, 169, 2993, 1, 935, 8305, 139],
 [62, 1, 317, 185, 8, 193, 69, 1049, 15, 73, 18, 592, 582],
 [16, 9, 519, 35, 217, 5899, 39, 1, 473],
 [345, 227, 95, 89, 128],
 [5900, 162, 28, 1430, 1430, 637, 1301, 510],
 [412, 503, 317, 1, 10, 8306],
 [75, 2994],

[37, 1, 7, 97, 245, 37, 722, 467],
 [289, 31, 398, 258, 289, 6],
 [259, 1, 2087, 258],
 [97, 2717, 899, 3362, 21, 184, 3363, 5901, 3364, 1686, 1158, 2718, 1678],
 [12, 408, 1, 256, 15, 48, 79, 60, 4],
 [4681, 3899, 1],
 [177, 5838, 1801, 340, 410, 146, 13, 1050],
 [204, 205, 625],
 [235, 38, 80, 10, 235, 156, 1, 63, 140, 383, 14],
 [8307, 368, 12, 3, 322, 45, 1307, 1005],
 [590, 5902, 165, 173, 545, 7],
 [1159, 1483, 311, 2719, 1416, 1483, 78, 2988, 51, 8],
 [179, 116, 191, 10, 116, 478, 1416, 1483, 236, 2719],
 [3900, 2478, 17, 1416, 126, 10, 116],
 [1483, 1483, 10, 116, 17, 1416, 1483, 1119, 236, 2988],
 [154, 115, 1, 73, 4, 68, 1493, 2995],
 [236, 2, 686, 1, 10, 116, 85, 2988, 9, 413, 1308],
 [32, 1956, 5, 537, 1375, 116, 8308, 29],
 [387, 8309, 360, 3901, 256, 381, 13, 46, 450, 578],
 [30, 993, 10, 116, 17, 4680, 1416, 1483, 171, 4682, 993, 1494],
 [90, 2995, 175, 4, 936, 1, 6, 8310],
 [2088, 168, 72, 99, 2267, 1, 17],
 [144, 284, 1309, 349, 353, 173],
 [683, 1431, 10, 63, 28, 5, 108, 1951],
 [8311, 2, 478, 1376, 1416, 17, 478, 1416, 1415, 1415],
 [569, 62, 244, 1, 30, 1051, 36, 1086],
 [16, 1, 20, 3902, 9, 375, 8312, 635, 176],
 [6, 308, 3, 133, 50, 81],
 [2996,
 110,
 85,
 2268,
 767,
 162,
 3903,
 464,
 87,
 8313,
 767,
 3904,
 155,
 472,
 5903,
 464,
 247,
 10,
 1],

[2720, 5904, 8314, 44, 7, 1, 2269],
 [535, 1, 17, 227],
 [47, 1, 12, 8315],
 [823, 1, 1, 533, 1],
 [25, 4, 1],
 [1052, 79, 3],
 [350, 15, 20, 7, 6, 1],
 [110, 17, 236, 478, 1416, 1308],
 [3905, 183, 1, 539, 11, 138],
 [1, 20, 44, 264, 727, 1298, 14, 14],
 [4683, 1485, 1485, 5905, 2997, 78, 345, 1],
 [20, 2078, 1957, 8316, 8317, 1081, 298, 3, 8318, 3906],
 [143, 1120],
 [3, 198],
 [8319, 1687, 1432, 1, 546, 1310, 118, 10],
 [129, 1, 1006, 1, 17, 1006],
 [5906, 583, 1, 8, 456, 1433, 798, 1311, 380, 1433, 584],
 [5, 40, 106, 1247, 274],
 [1, 113],
 [1, 20, 143, 20, 593],
 [1, 362, 372],
 [143, 4, 322, 141, 1, 133, 468, 141, 97, 50, 44],
 [359, 285, 4, 89, 199, 214, 145, 378],
 [4684,
 1958,
 71,
 8320,
 2089,
 158,
 379,
 3907,
 2479,
 8321,
 19,
 4685,
 1802,
 2998,
 59,
 5,
 78,
 3365],
 [4686, 3, 8322],
 [5907, 8323, 55, 1],
 [204, 204, 17, 421],
 [30, 27, 1, 2270, 228, 3908, 11, 5908, 90, 50, 1495, 575],
 [110, 1046, 1],
 [5909, 207, 2, 117, 1, 25, 4, 48, 169, 511, 11],

[1432, 8324, 8325, 8326, 9, 15, 487, 8327, 1053, 5910],
 [439, 26, 823, 1496, 439, 8328, 1],
 [366, 3, 371, 198, 8329, 366, 371, 8330, 553, 3],
 [15, 100, 1, 15, 180, 6],
 [23, 23, 570, 263, 6, 4, 8331, 24, 162, 87, 356, 208, 2721],
 [865, 6, 37, 865, 6, 37, 66, 6, 95, 22, 1],
 [90, 80, 5, 274, 106, 842, 1087],
 [965, 2262, 69, 1, 208, 146, 766, 3366, 1424],
 [92, 1, 92, 99],
 [85, 5, 4, 44, 6, 648],
 [3909, 18, 961],
 [306, 25, 81, 1, 154],
 [8332, 182, 3, 7, 25, 4],
 [74, 27, 1, 6, 425, 21],
 [1434, 281, 332, 28, 1, 565],
 [1047, 6, 539, 30, 661, 1, 539, 608, 133, 56, 37, 277, 456],
 [10, 23, 1294, 200, 1688, 343, 1, 54, 4687],
 [8333, 5911, 846, 8334, 37, 2999, 255, 8335, 388, 5911, 4688, 24],
 [8, 11, 1, 427, 1088],
 [8, 7, 38, 2090, 36, 95, 2271],
 [19, 3910, 368, 125, 22, 540, 3911, 415, 17, 3],
 [15, 97, 3912, 1, 222],
 [192, 167, 192, 3, 236, 51, 683],
 [1, 1036, 3],
 [1048, 49, 93, 701, 547, 8336],
 [56, 6, 4689, 1367, 782, 353, 465, 229, 7, 1, 28, 44, 11],
 [309, 900, 3, 900, 3, 195, 3, 3, 241, 333, 3, 2480, 3000, 302],
 [267, 18, 25, 349, 28, 474, 5847, 19, 3913, 65],
 [267, 5, 35],
 [1054, 470, 443, 4690, 1089, 1689, 167],
 [2270, 1],
 [9, 101, 3, 58, 191, 78, 130, 101, 3, 58, 191, 78, 130],
 [164, 1, 1690],
 [15, 647, 413, 10, 574, 1, 54, 4, 481, 16, 150, 99, 25, 647, 271],
 [4691, 3914, 609, 415, 4692, 702, 3001],
 [54, 1],
 [802, 1959, 1, 802, 8337, 1691, 2084],
 [1052, 144, 1052, 1, 5912, 8338],
 [17, 3, 143, 763],
 [70, 46, 30, 290, 8339, 1, 172, 74, 290],
 [70, 3],
 [474, 72, 110, 783, 454, 329, 1377],
 [21, 34, 1, 117, 56, 748, 136, 1007, 666],
 [12, 1, 557, 4, 3367, 8340, 966, 4693, 1160, 542, 3368, 1, 8341],
 [12, 592, 177, 59, 318, 93, 1],
 [288, 142, 1, 317, 5913, 6],
 [1, 1, 1, 276, 8342],

[1048, 35, 56, 1592, 223, 1, 8343, 303],
 [435, 178, 1196, 28, 1],
 [232, 8, 125, 310, 54, 1, 1, 328, 15, 151, 8344, 8345, 2481],
 [12, 397, 155, 1, 88],
 [15, 11, 1692, 8346, 346, 158, 19, 594, 1312, 213, 15, 31, 585],
 [1497, 323, 69, 1803, 896, 825, 9, 1],
 [626, 315, 1, 1498, 299, 22, 299, 299],
 [50, 4, 485, 4, 2722, 5, 46, 15, 485, 4],
 [1302, 85, 5, 234, 237, 1302, 85, 5, 237, 151, 421],
 [344, 504, 4, 538, 132, 42, 116, 45, 125, 5, 42, 197, 41, 566],
 [260, 718, 2091, 399, 207, 55, 2723, 260, 718, 1, 399, 207, 2092],
 [10, 9, 9, 350, 9, 5914, 1, 166, 3915, 3916, 3917, 8347, 5915, 256, 2724],
 [901, 4, 4694, 1, 18, 38, 447],
 [51, 407, 54, 1, 1593, 2725],
 [172, 5, 221],
 [826],
 [1594, 7, 1693, 937, 205, 8348],
 [8349, 3918, 5],
 [5916, 72, 1248, 3369, 72, 136, 1, 17, 1960, 8350, 897, 1694, 223, 8351],
 [1495, 4, 8352, 728, 1, 4695, 339, 130],
 [73, 902, 8353, 59, 2093, 3919, 13, 277, 14],
 [303, 369, 1, 844, 3920, 2726, 967],
 [401, 11, 28, 404, 1, 7, 1435, 191, 2, 860, 3921, 890, 338, 156, 440, 128],
 [246, 158, 22, 1, 42, 3370, 6],
 [5917, 5918, 1, 3922],
 [17, 1008, 22, 1, 106, 1804],
 [60, 38, 108, 5],
 [27, 1, 91, 299, 193, 74, 19, 512, 169, 4696, 299, 131],
 [465, 1695, 1, 3371, 7, 166, 287, 243, 8354],
 [336, 68, 92, 3, 2727],
 [1120, 758, 439, 544],
 [61, 412, 8355, 1436, 1805, 15, 1, 265, 4, 23, 88, 239],
 [8356, 9, 4697, 9, 37, 199, 5919, 1437, 1, 15, 103, 2272, 1161, 9, 8357, 92],
 [41, 729, 1313, 1],
 [278, 2273, 278, 278],
 [3372, 2482, 121],
 [324, 94, 2728, 97],
 [8358, 5, 29, 638, 4, 288, 68, 18, 638],
 [14, 961, 4698, 1961, 2, 3002, 3373, 281, 2274, 822, 3923],
 [510, 5920, 3924, 337],
 [190, 220, 143, 245, 419, 381],
 [4, 2275, 217, 3003, 5921, 4699, 1, 79],
 [47, 2, 860, 1499, 1009, 25, 4, 5],
 [47, 4700, 3374, 188, 11, 96],
 [8359,
 8360,
 8361,

8362,
 5922,
 3375,
 5923,
 5924,
 805,
 5924,
 1378,
 67,
 5925,
 8363,
 5926,
 2094,
 4701,
 3,
 604,
 1500,
 317,
 4702,
 8364],
 [337, 20, 119, 337, 109, 1, 41, 15, 23, 1113],
 [432, 126, 3925, 2729, 3376, 128, 124, 2483, 89, 8, 4, 4703, 172, 163],
 [900, 344, 54, 1696, 160, 33, 52, 32, 1790, 1697, 956, 65, 3377],
 [600, 177, 57, 336, 4704, 63, 4705, 95, 555, 87, 87, 3378, 1],
 [510, 8365, 368, 5927, 1438],
 [97, 3877, 5, 1249, 50, 20, 202, 4, 41, 50],
 [211, 19, 181, 2276, 23, 842, 268, 4706, 2095, 264, 38, 4, 63, 1, 42],
 [211, 22, 194, 162, 8366, 136, 31, 8367, 34, 112],
 [72, 3379, 208, 1010, 14, 1, 784, 307, 4],
 [6, 76, 1694, 10, 1, 8, 79, 2730, 180, 35, 20, 1],
 [6, 1685, 5, 8368, 569],
 [1, 5928],
 [1501, 3, 222, 312, 195],
 [5929, 2484, 2731, 35, 13],
 [301, 420, 7, 246, 3, 98, 41, 12, 79, 46, 464, 180],
 [408, 248, 993, 2277, 633, 150, 1502, 1090, 779, 337, 1053, 30, 1371, 86],
 [43, 33, 5930, 649, 87, 350, 26, 8369, 189, 250, 89, 5931, 8370],
 [2732, 1379, 8371, 463, 22, 1, 347, 473, 3926, 3926, 368],
 [2, 222, 312, 2485, 5, 52, 1003, 1439],
 [4707, 8372, 443, 434, 36, 2278, 1],
 [8373, 16, 391, 3],
 [246,
 5932,
 265,
 42,
 1698,
 52,

38,
 301,
 968,
 1698,
 5933,
 349,
 331,
 749,
 1311,
 253,
 21,
 1],
 [3927,
 2486,
 580,
 162,
 49,
 5934,
 173,
 8374,
 2279,
 23,
 8375,
 8376,
 627,
 261,
 1572],
 [969, 724, 2487, 215, 750, 213, 1],
 [1147, 1088, 1055, 166, 261, 223, 357, 1088],
 [2280, 3380, 2733, 703, 40, 703, 40, 1],
 [68, 3004, 5, 619, 104, 308, 5935, 3928, 5],
 [50, 1250, 28, 456, 61],
 [1, 743],
 [1162, 12, 2470, 819, 5936, 5],
 [768, 10, 1, 15, 263, 1503, 269, 310],
 [29, 970, 610, 35, 202, 4, 242],
 [59, 5, 253, 2734, 3381, 415],
 [59, 178, 9, 18, 85, 5, 8377, 10],
 [1582, 243, 1440, 628, 5, 15, 91, 53],
 [191, 385, 3, 385],
 [19, 3382, 178, 5937, 100, 1121, 267, 462, 12, 311, 804, 309, 1248],
 [1247, 4708, 927, 415, 1806, 4709, 49, 392, 94, 114, 806, 8378],
 [464, 94, 3888, 1314],
 [3, 198],
 [3, 106, 1163, 3929],
 [2735, 967, 926, 1, 95, 55, 130],
 [31, 5938, 1807, 10, 12, 4710, 364, 1],

[89, 7, 938, 8, 8379],
 [4711, 13, 2281, 181, 320, 3930, 1186, 201],
 [265, 5939, 1, 6, 2246, 123, 3, 3910, 25, 638],
 [89, 1595, 5940, 8380],
 [4712, 54, 3005, 204, 179, 3005, 204, 204, 2488, 338],
 [469, 1197, 3931, 49, 7, 60, 13, 69, 405, 11, 114],
 [225, 8381, 1],
 [97, 97, 1, 97, 62],
 [166, 1251, 8382, 94, 1673, 2259, 16, 8383, 2482, 8384, 903],
 [9, 12, 8385, 1, 6, 847, 8386, 3383, 8387],
 [2736, 415, 2736, 956, 94, 5941, 51, 4, 375, 133, 22, 3006, 379, 129, 9, 125],
 [344, 352, 12, 3, 428, 29, 19, 95, 513],
 [1699, 5, 1091, 5, 1154, 5],
 [8388, 668, 1700, 9, 118, 1, 133, 12, 107, 1252, 751, 3932, 386],
 [1109, 1, 4, 6, 129],
 [34, 65, 34, 31, 34, 1092, 3933, 2467],
 [300, 87, 247, 416, 4713, 4, 745],
 [124, 1, 4714, 50],
 [68, 1427, 2282, 2, 860, 310, 2, 147, 86, 81, 5],
 [39, 209, 3, 458, 1504, 646],
 [10, 1163, 2737, 6],
 [43,
 32,
 183,
 1441,
 368,
 2096,
 246,
 20,
 862,
 253,
 704,
 87,
 77,
 650,
 468,
 1939,
 299],
 [9, 64, 1, 785, 2283, 1, 8389, 195],
 [51, 20, 126, 5, 357],
 [51, 56, 705, 3934, 1505, 1788, 1],
 [51, 20, 5, 141, 1506, 185, 29, 752, 4692],
 [8390, 1, 8391],
 [138, 8, 3384, 40, 563],
 [381, 1, 197, 224, 8392, 266, 378, 26],
 [566, 1934, 72, 143, 19, 207, 1, 1093, 3007],
 [1307, 50, 199, 5942, 43, 8393, 5943, 5944],

[8394, 8395, 2489],
 [22, 1, 3008, 152, 463, 904, 564, 5945, 1239],
 [22, 8396, 5],
 [22, 577, 429, 4, 3385, 5946, 4, 2284, 1, 353],
 [22, 638, 1, 2, 16, 1233, 2097, 409],
 [1507, 11, 2490, 3009],
 [1, 356, 1701, 241, 235],
 [22, 88, 5, 406],
 [71],
 [316, 5, 3337, 5947, 1086, 1087, 8397, 2738],
 [61, 140, 89, 61, 89],
 [4, 9, 159, 470, 159, 49, 10],
 [12, 520, 3],
 [105, 354, 333, 8398, 2491, 87, 389, 38, 196, 3386, 75],
 [25, 216, 120, 42, 8399, 3935, 3],
 [25, 4, 64, 24, 494, 8, 1421, 389],
 [667, 5948, 329, 7, 1, 8400],
 [60, 4, 77, 96, 289, 8401, 8402, 786, 1315, 62],
 [326, 1572, 135, 23, 96, 3936, 259, 214, 2492, 289],
 [2285, 897, 1, 314],
 [158, 11, 92, 7, 71, 18, 530, 290, 1011, 7, 751, 656, 466],
 [300],
 [1948, 178, 27, 487, 1122, 3],
 [821, 4, 753, 2984, 312, 116, 94, 611, 274],
 [1940, 10, 130, 294, 4, 13, 2, 860, 3937, 56, 126, 14],
 [33, 3, 8, 14, 195, 26, 9],
 [1],
 [2286, 46, 190, 3387, 5949, 1, 25, 3938, 44, 730, 37, 3010, 2283],
 [1, 1123, 11],
 [1, 1369, 64, 8403],
 [17, 211, 1, 17],
 [5, 2098, 1962, 8404, 91, 147, 638],
 [54, 8405, 2287, 69, 39, 3939, 5950, 3],
 [1675, 586, 3, 535, 3917],
 [40, 5951, 231, 73, 178, 107, 4715, 348, 10, 8406, 5952, 1],
 [48, 141, 1, 1508, 228, 8407, 8408],
 [70, 167, 1, 540, 25, 4, 8409, 8410],
 [10, 5, 7, 1124, 18, 200, 53, 317, 531, 140, 1808, 968, 32, 29, 119],
 [997, 8411, 8412, 8413, 2739, 8414],
 [3, 200, 53, 905, 25, 4, 889, 8415, 105, 130, 55, 5953, 3388, 8416],
 [1963,
 138,
 370,
 671,
 2493,
 659,
 21,

45,
 80,
 5954,
 247,
 3389,
 457,
 423,
 5955,
 157,
 10,
 1],
 [200, 53, 1589, 3011, 412, 261, 93, 439],
 [39, 73, 1164, 1, 1509, 1164, 1509, 14],
 [3390, 1, 3391, 8417],
 [6, 1, 1, 7, 6, 2092, 7, 3, 114, 7, 2288, 7, 3, 77, 1313, 1416, 8418],
 [5,
 152,
 18,
 2740,
 54,
 3940,
 848,
 4,
 170,
 22,
 199,
 170,
 963,
 22,
 11,
 170,
 199],
 [132, 36, 3, 32, 216, 639, 842, 406, 2741, 2289, 334, 5956, 8419],
 [353,
 1,
 21,
 109,
 206,
 930,
 930,
 930,
 287,
 579,
 353,
 8420,
 359,
 103,

139,
 119,
 206],
 [110, 10, 1, 151, 1198, 747, 649],
 [706, 849],
 [1160, 971, 35, 225, 3],
 [508, 83, 3941, 3392, 28, 704, 3012, 71, 8421, 87, 532, 1011, 34, 9],
 [574, 29, 16, 5957, 5958, 864, 112, 3],
 [102,
 26,
 3942,
 113,
 845,
 1702,
 178,
 19,
 1,
 1964,
 4,
 1809,
 3393,
 429,
 4,
 5959],
 [824, 26, 488, 23, 111, 623],
 [81, 1810, 5, 47, 3013, 46, 113],
 [21, 5, 46, 172, 53, 887, 1808, 968],
 [71, 45],
 [253, 1, 906, 1125, 566, 213, 932],
 [3943, 3, 469, 3, 1965, 3, 5960, 3, 92, 3, 178, 17, 827, 3],
 [8, 105, 1, 2494, 20],
 [869, 2495, 662, 4, 1, 3944, 2742, 2743, 4716, 966],
 [2, 117, 376, 75, 1966, 5961],
 [5, 111],
 [5, 1],
 [226, 1125, 27, 459, 1703, 274, 939, 405, 49, 13, 475],
 [8, 749, 2496, 279, 46, 8, 1, 235, 38],
 [1, 476, 5854],
 [35, 318, 2099, 8422, 124, 35, 124],
 [5],
 [3945, 75, 2100, 3946, 8423],
 [5962, 1, 1, 53, 1199, 8424],
 [2290, 54, 2101, 1811, 649, 75, 1012, 201, 263, 4717, 5963, 8425, 1812],
 [36, 1, 5964, 1967, 4718],
 [2291, 44, 1, 110, 110, 2706, 669, 4719, 962, 38, 1813],
 [1, 17, 2084],
 [129, 183, 408, 197, 1596, 250, 24, 51, 8426, 268, 3394, 51, 718, 8427],

[4720, 1442, 3395, 529, 454, 331, 348, 86, 1],
 [1968, 166, 360, 449, 972, 3014],
 [1968, 34, 16, 8428, 2744, 1380, 8429, 16, 57, 640],
 [147, 568, 577],
 [5965, 25, 24, 1056, 338, 8430, 1381, 1057, 117, 4721],
 [148, 8431, 2497, 340, 4722, 1704, 1814, 8432],
 [4723, 3932, 121, 1815, 656],
 [8433, 463, 1200, 437, 542, 102, 383, 4, 1],
 [291, 49, 13],
 [291, 49, 13, 2998],
 [116, 15, 111, 1000, 188, 514, 2292, 4],
 [1705, 7, 219, 170, 2745, 228, 21, 357],
 [8434, 89],
 [1, 156, 309, 320, 8435, 277],
 [1379, 220],
 [13, 1316, 263],
 [4724, 5966, 1443, 1, 12, 28, 1201],
 [2102, 110, 8, 5, 171, 8436, 28, 551, 139, 70, 3333, 1126, 103, 515, 431],
 [95, 131, 12, 2746, 702, 154, 257, 7, 235],
 [907, 1376, 1058, 184, 908, 2471, 593, 4725, 1969, 133, 12, 28, 53],
 [1, 360, 38, 199, 3396, 752, 3015, 8437, 8438, 4726],
 [15,
 199,
 3016,
 5967,
 8439,
 265,
 3947,
 121,
 4727,
 5968,
 686,
 4728,
 2293,
 3017,
 3018,
 8440,
 419,
 691],
 [2470, 8441, 343, 8442, 133, 1039, 360, 3397, 909, 787, 940, 609, 3948],
 [5969, 28, 339],
 [8443, 49, 13],
 [43, 263, 451, 49, 13],
 [3949, 495],
 [309, 194, 4729, 24, 1706, 27, 42, 12, 33],
 [309, 5970, 3019, 85, 5971, 8444, 3950, 2747],
 [88, 160, 2498, 4708, 5972, 8445, 338, 123, 1165],

[2, 1080, 139, 29, 696, 5, 10],
 [175, 122, 3398, 40, 321, 3, 30, 523],
 [973, 1059, 3951],
 [2098, 1597, 3020],
 [3021, 905, 1376, 1, 7, 10, 32, 969, 519, 289, 4730, 3952],
 [1, 83, 155, 910, 287],
 [3, 8, 8446],
 [8447, 195, 556, 361, 204, 50, 312, 50, 8448],
 [822, 2294, 3953, 8449, 244, 731, 1035],
 [51, 41, 194, 1681],
 [34, 285, 8450, 226, 973],
 [16, 27, 1, 56, 807, 5973],
 [2, 2, 148, 570, 2499, 3399, 240],
 [1816, 3400, 244, 3401, 8451, 4731, 8452, 204],
 [4732, 74, 370, 568, 577, 2280, 4733, 2103, 1970, 4734, 50, 4734, 8453],
 [728, 149, 3, 344, 54, 8454, 3402],
 [469, 3],
 [3, 173, 409, 541, 4735, 1598],
 [2075,
 4736,
 5,
 44,
 2295,
 4,
 4737,
 3022,
 2104,
 8455,
 1158,
 373,
 1599,
 219,
 18,
 207],
 [507, 243, 1238, 1363, 1817, 381],
 [23, 8456, 1314, 671, 110, 2500, 1707, 1314, 970, 907],
 [139, 3954, 8457, 90, 267, 12, 5, 180, 45, 47, 1060],
 [85, 66, 5, 81, 1818],
 [118, 179, 938, 31, 384, 170, 769, 870, 2105, 732, 871, 187, 938],
 [45, 226, 45, 5974],
 [176, 95, 96, 2296, 2273, 3023, 17, 5975, 8458, 1039, 3403],
 [291, 49, 13],
 [1510, 3955, 3404, 5976, 2748, 8459, 5977, 1819, 2297, 372, 651, 5978],
 [243, 524],
 [14, 1684, 4, 2, 309, 99, 1, 2749],
 [3024, 233, 2106, 60, 4],
 [3, 1148, 27, 46, 57, 1, 1820, 9, 1253],

[9, 81, 995, 661, 1, 436, 80, 9, 79, 8460, 2074, 1821, 2750],
 [291, 49, 13],
 [1444, 1, 200, 53, 1254, 143, 866, 224, 427, 2298, 2501],
 [2502, 9, 911, 911, 78, 3956, 35, 8, 8461, 8462, 8463, 43, 32, 300],
 [4738, 1511],
 [259, 2503, 162, 3405, 24, 764, 893, 95],
 [368, 83, 26, 5979, 3025, 1822],
 [3406, 3953, 3957, 1968, 607, 2107, 1708, 3958],
 [309, 49, 5980, 1202, 822],
 [3026, 2273, 5980, 1202, 822],
 [529, 2108, 872, 1708, 107, 172],
 [116, 539, 5, 1166, 6, 1127, 888],
 [1, 576, 2751, 4739, 4679, 2109, 3407, 373, 452, 33, 650, 9, 35, 8, 524],
 [3959, 4740, 8, 444, 24, 1445],
 [1251, 582, 110, 1971, 2110],
 [4741, 5981, 744, 5982, 1, 521],
 [64, 491, 34, 1249, 15, 35, 84, 1, 262, 10],
 [2, 1, 181, 2504, 1057, 55, 1234, 888],
 [2468, 3, 4742, 8464, 145, 210, 315, 567, 41, 1255, 450],
 [8465, 8466, 490, 1701, 283, 4743, 3960, 44, 3027, 4744, 5983, 5984],
 [1],
 [35, 1, 12, 16, 217, 93, 265, 8467],
 [9, 417, 242, 2111, 66, 1512],
 [3, 59, 5985, 733, 53, 1709, 108, 456],
 [1681, 4745, 4, 95, 8468, 253, 1],
 [590, 1972, 93, 8469, 8470, 3028, 1317, 8471, 145],
 [1709, 1600, 134, 115, 4746, 115, 14],
 [8472, 8473, 762, 63, 1, 8474, 107],
 [3961, 402, 3408, 8475, 8476, 972, 492],
 [5986, 48, 3962, 5987, 69, 102, 320, 496, 127, 691, 5988, 3962],
 [75, 8477],
 [1039, 966, 5974, 158, 1086],
 [231, 324, 161],
 [233],
 [233],
 [3029, 4747, 233],
 [2, 707, 558, 3, 101, 8],
 [13],
 [2, 8478, 65, 56, 10, 122, 472, 2299, 3409, 40],
 [544, 114, 5989, 183, 96, 422],
 [291, 49, 13],
 [5, 78, 3030, 400, 57, 78],
 [2,
 863,
 1203,
 1513,
 4748,

3963,
 3410,
 2752,
 1446,
 2300,
 115,
 5990,
 357,
 24,
 474,
 3031,
 3963],
 [17, 763, 1823, 1823],
 [161],
 [2, 2753, 75, 392],
 [253, 1, 906, 799, 8, 447, 5991, 5992, 33, 902],
 [253, 1, 906, 189, 1708, 356, 208, 423, 208],
 [253, 1, 906, 650, 196, 964, 8479, 64, 988, 1447],
 [315, 213, 932, 1125, 213, 25, 4, 3032, 4749, 817, 1],
 [43, 41, 148, 548, 33],
 [126, 1601, 6, 78, 1, 67, 2078],
 [80,
 66,
 53,
 9,
 68,
 355,
 3964,
 66,
 5,
 328,
 9,
 1602,
 53,
 67,
 1,
 181,
 727,
 3965,
 8480],
 [84, 1, 324],
 [10, 298, 3, 81, 17, 298, 3],
 [5993, 2505, 1973, 7, 220, 405, 781],
 [8481, 8482, 6, 6, 6, 6, 1092, 1092, 1092, 1092, 6, 6, 6, 6],
 [129, 155, 1],
 [670, 161],
 [8, 264, 8483, 3411, 419, 4, 4750, 4751, 241, 220],

[291, 49, 13],
 [3016, 8484, 2754, 8485, 171, 454, 1318, 651, 3033, 8486, 4752],
 [291, 49, 13],
 [3034, 1710, 240],
 [3, 3035, 33],
 [23, 1, 445, 1603, 327, 1604, 162, 331],
 [2273, 3412, 535, 248, 993, 610, 2301, 825, 61, 8487, 2755],
 [],
 [20, 4753, 38],
 [48, 126, 1117, 2506, 3966],
 [3967, 102, 3961, 3968, 3036],
 [155, 602, 5994, 8488, 8489, 1605, 4754],
 [3969, 243, 105, 1448, 142],
 [63, 8490, 8491, 127, 86, 3970, 360, 1256, 5995, 1974],
 [63, 124, 754, 385, 5996, 42, 2711],
 [1449, 1597, 320, 496, 86, 788, 1319, 415, 847, 503, 317, 2302, 1689, 3413],
 [260,
 867,
 3022,
 3971,
 285,
 2507,
 3414,
 334,
 3415,
 2756,
 1085,
 1711,
 4755,
 26,
 199],
 [1597,
 691,
 1606,
 20,
 308,
 277,
 315,
 3972,
 3416,
 708,
 222,
 18,
 1579,
 754,
 1122],
 [311, 2112, 38, 3417, 68, 106, 83],

[3973, 1975, 416, 788, 309, 126],
 [5997, 12, 207, 1382, 4756, 799, 1257, 46, 81, 33, 8, 502, 5998, 54, 33],
 [12, 8492, 781],
 [61, 912, 86, 5999, 1163],
 [3035, 501],
 [169, 2113, 95, 107, 190, 308, 2091, 24, 16, 18],
 [1081, 1, 6000, 153, 323, 1976],
 [672, 49, 13],
 [2, 4757, 16, 393, 8493, 1, 42, 104],
 [652, 4758, 998, 309, 828, 378, 1607, 828, 1],
 [],
 [1],
 [1, 8494, 2114, 3037, 1712],
 [1450, 3, 171],
 [8495, 1451],
 [524, 1320, 574, 18, 2508, 1128],
 [8496, 8497, 145, 217, 2503, 3915, 8498],
 [264, 49, 13],
 [234, 274, 1167, 60, 682, 1713, 8499, 47, 1714, 291, 2757],
 [3418],
 [17, 3],
 [1],
 [265, 3974, 8500, 8501, 115, 1054],
 [482, 1364, 6001, 1061, 18, 6002, 8502, 113, 79, 525],
 [6003, 8503, 1, 4759, 251, 29, 99, 8504, 8505],
 [35, 1452, 30, 320, 1195, 1243],
 [162, 8506, 205],
 [1, 1040, 2485],
 [182, 12, 1114, 1715, 291],
 [1168],
 [],
 [3975,
 217,
 8507,
 873,
 214,
 371,
 2509,
 214,
 1977,
 23,
 1716,
 371,
 3038,
 939,
 4760],
 [],

[],
 [8508],
 [1, 8],
 [2758, 1, 5, 5, 6004, 551, 2303, 2510],
 [822, 2294, 8509, 3419, 3953, 87, 6005, 2290, 4761, 388],
 [8, 2261, 536, 66, 1, 105],
 [6, 1, 104, 2, 88, 483, 165, 1, 129, 9, 6006, 666, 130, 78, 66, 329],
 [81, 103, 308, 277],
 [2759, 6007, 550, 126, 1086, 8510, 653],
 [6008, 2759, 8511, 1],
 [1796, 550, 7, 158],
 [2996, 2115, 300, 2759, 6009],
 [3039, 4762, 829, 8512, 5, 7],
 [3976, 829, 5],
 [789, 19, 3420, 5, 500, 2511],
 [216, 8513, 466, 107, 941, 126, 8514, 5],
 [156, 5, 730],
 [4763, 5, 3977, 673],
 [3978, 5, 610, 2760, 730],
 [8515, 2304, 5, 8516],
 [5, 10, 730],
 [269, 1453, 19, 909, 71, 8517],
 [3421, 1608, 3979, 6010, 200, 3422],
 [3979],
 [3979],
 [3976, 872, 4762, 107, 641, 5],
 [8518, 913, 941, 5, 21],
 [1796, 550, 584, 8, 6011, 8519],
 [144, 2305, 351, 107, 5],
 [2697, 285, 6012, 1383, 5],
 [1166, 642, 6012],
 [1204, 115, 123, 909, 5, 3980, 628],
 [909, 8520, 2734, 2761, 909, 248, 5],
 [909, 5],
 [909, 5],
 [299, 264, 1113, 3981, 5],
 [3978, 88, 21, 1, 8521],
 [8522, 5, 434, 8523, 658],
 [207, 129, 351, 3420, 5],
 [590, 6013, 1, 6014, 207, 129, 3978],
 [143, 5],
 [227, 269, 1453, 19, 236, 71, 158, 658],
 [269, 1453, 19, 909, 71, 8524],
 [8525, 352, 5],
 [2306, 281, 8, 5, 8526],
 [2306, 1803, 107, 5],
 [2306, 842, 5],

[1818, 19, 276, 5],
 [1, 2762, 108, 476, 71, 4764],
 [236, 114, 550, 8527],
 [123, 1, 4765, 269, 2254],
 [3982, 2760, 108, 269, 123, 770, 5],
 [850, 941, 89, 584, 21, 8528],
 [6015, 1818, 1205, 1824, 826],
 [909, 8529, 2763, 8530, 4766, 909, 5],
 [2307, 5, 276],
 [2, 1978, 16, 51, 117, 2500, 91, 24],
 [2, 2, 2, 2, 57, 1672, 3, 148],
 [3, 168, 8, 502, 140, 908, 266, 696, 1485],
 [6016, 1979],
 [4767, 8531, 75, 392],
 [4768, 2308, 1825, 1045, 830, 153, 4769],
 [51, 134, 8532, 48, 831, 60, 1454, 4770, 24],
 [27, 1],
 [3423, 1321, 143],
 [344, 29, 3424, 210, 1],
 [326, 8533, 160, 1826],
 [475, 4, 3425, 4771, 3983, 2764, 593, 3984, 771, 3426, 8534, 138, 227, 117],
 [1, 1094, 1322, 328, 1944],
 [3384, 283, 1502, 4678, 8535, 8536, 3985],
 [1],
 [211, 73, 404, 73, 360, 1],
 [1, 8, 778, 138, 331, 33, 331, 422],
 [1980, 167, 389, 785, 8537, 1481, 8538, 23, 914, 60],
 [],
 [1],
 [27, 1],
 [1],
 [1609, 52, 1981, 1981, 1960, 6017, 1692, 5],
 [401, 102, 1384, 4, 5, 674, 292, 323],
 [9, 44, 1],
 [548, 3427, 18, 430, 6018, 8539, 132, 8540, 253, 1],
 [155,
 1385,
 8541,
 247,
 124,
 434,
 669,
 3040,
 1717,
 487,
 274,
 974,

652,
 1827,
 71,
 52,
 1514,
 8542,
 1206],
 [1828, 8543, 8544, 69, 346, 193, 372, 16, 6, 6019, 3, 6020, 6021],
 [2, 1515, 333, 3428, 1516, 965, 1515, 13, 60, 333, 6022],
 [2, 2, 696, 5],
 [2, 17, 1718, 2, 743, 10, 3, 85, 1718],
 [1717, 4772, 790],
 [2, 74, 1, 78, 6023, 780, 4773, 158, 462, 16, 33],
 [412, 444, 24, 8, 991, 150],
 [413, 584, 17, 3, 17, 702, 348],
 [2, 3, 12, 149, 564, 733, 53],
 [2, 1982, 1258, 6, 97, 42, 101, 4, 1],
 [2, 1, 552, 1829, 305, 10, 72, 2116],
 [1, 487, 38, 2512],
 [41, 342, 2309, 6024, 342, 104, 41, 162, 342, 11, 104],
 [2, 6, 6025, 3, 3, 1830, 3986, 35, 1610, 8545],
 [2, 348, 65, 219, 1808, 1149, 656],
 [2, 4, 708, 1, 657, 80, 10],
 [2, 407, 146, 65, 547, 595, 2513],
 [2, 286, 1, 3987, 3429, 199, 1719, 4, 966, 635],
 [],
 [2,
 165,
 8546,
 181,
 423,
 357,
 1306,
 434,
 8547,
 3041,
 229,
 153,
 233,
 181,
 612,
 165],
 [2117, 791, 3, 1720, 8, 942, 102, 169, 1095, 105, 150],
 [12, 103, 1, 987, 4641, 6026, 1079],
 [353, 8548, 23, 612, 214, 187, 101, 11, 2118, 935, 216, 1],
 [1386, 59, 733, 53, 3],
 [2, 353, 84, 165, 1387, 3430, 1, 324, 231, 157, 22, 7],

```

[2, 832, 3, 2, 52],
[47, 6, 18, 355, 314, 204, 779, 221, 179, 1373, 8549],
[121, 8, 874, 8550],
[833, 1, 17, 2310, 3988, 284, 216, 8551, 237, 520, 596],
[5, 503, 53, 4, 6027],
[5],
[2, 198, 5],
[1],
[1517, 99, 3989, 643, 332, 869, 24, 190],
[405, 492, 10, 9, 1693],
[994, 165, 1169, 11, 14, 311, 1, 98, 692, 3431],
[2, 1831, 123, 3432, 751, 223, 1, 936],
[2, 297, 2, 348, 904, 1422, 22, 820, 235, 9, 167, 151, 8552, 31],
[186, 3, 68, 109],
[394, 73, 317, 2, 2070, 1, 8553, 3990, 2119, 122],
[808, 223, 809, 686, 141, 1, 136, 1983, 38],
[851, 500, 13, 62, 3991, 908],
[1611, 1813, 4, 3992, 2, 6, 4774, 1],
[1, 8, 989, 209],
[670, 161, 497, 2120, 3890],
[8554, 232, 2, 8, 2121, 889, 1612, 178, 75, 4775],
[1322, 54],
[6, 1832, 4776, 1, 4, 59, 693, 931],
[49, 1, 32, 1361, 1793],
[192, 526, 7, 1, 193, 301, 49],
[161, 68, 277, 291, 8555, 3],
[2, 3, 709, 41, 80, 594, 95, 6028, 1518, 7, 3, 10],
[2, 278, 8, 420, 1833],
[2, 1, 15, 12, 86, 188],
[1, 4777],
[1834, 5, 619],
[50, 2, 662, 68, 398, 227, 136, 1613, 8556, 3, 8557],
[211, 542, 1236, 1, 6, 1236],
[769, 6, 128, 1, 2122, 132, 170],
[1, 18, 274, 3042, 8558, 6029],
[1, 125, 330],
[34, 2, 1793, 121, 3433, 466, 108, 1122, 430, 809, 809, 809],
[2, 212, 2123, 5, 1829, 354, 6030],
[2, 3334, 733, 1],
[1, 4778, 30, 48, 15, 45],
[470, 27, 60, 13],
[3434, 1, 9, 271, 28],
[2, 128, 1, 3435, 7],
...]
```

```

[183]: tweet_data = pad_sequences(sequences,maxlen=max_len)
       tweet_data
```

```
[183]: array([[ 0,  0,  0, ..., 76,  61,  13],
             [ 0,  0,  0, ...,  3, 695, 370],
             [ 0,  0,  0, ..., 239, 925,  11],
             ...,
             [ 0,  0,  0, ..., 98,  95, 230],
             [ 0,  0,  0, ...,  1, 1660, 332],
             [ 0,  0,  0, ..., 94,  49, 16840]])
```

```
[184]: tweet_data.shape
```

```
[184]: (24782, 138)
```

0.3.2 Model Building

```
[185]: vocab_size = len(tokenizer.word_index) + 1
       vocab_size
```

```
[185]: 16841
```

```
[186]: cnn_representation_dim = 100
```

```
[187]: model = Sequential()
       model.add(Embedding(vocab_size, cnn_representation_dim, input_length=max_len))
       model.add(Conv1D(128, 5, activation='relu'))
       model.add(GlobalMaxPooling1D())
       model.add(Dense(128, activation='relu'))
       model.add(Dense(3, activation='softmax'))
```

```
[188]: model.summary()
```

Model: "sequential_5"

Layer (type)	Output Shape	Param #
embedding_2 (Embedding)	(None, 138, 100)	1684100
conv1d_2 (Conv1D)	(None, 134, 128)	64128
global_max_pooling1d_2 (GlobalMaxPooling1D)	(None, 128)	0
dense_3 (Dense)	(None, 128)	16512
dense_4 (Dense)	(None, 3)	387

Total params: 1765127 (6.73 MB)

Trainable params: 1765127 (6.73 MB)
Non-trainable params: 0 (0.00 Byte)

```
[189]: model.compile(optimizer='adam', loss='sparse_categorical_crossentropy',  
    ↪ metrics=['accuracy'])
```

0.3.3 Training and Get Representation for Down Stream Task

```
[192]: model.fit(tweet_data, labels, epochs=10, batch_size=32, validation_split=0.2)
```

```
Epoch 1/10  
620/620 [=====] - 11s 17ms/step - loss: 0.3427 -  
accuracy: 0.8753 - val_loss: 0.2419 - val_accuracy: 0.9179  
Epoch 2/10  
620/620 [=====] - 11s 17ms/step - loss: 0.1918 -  
accuracy: 0.9320 - val_loss: 0.2623 - val_accuracy: 0.9120  
Epoch 3/10  
620/620 [=====] - 10s 17ms/step - loss: 0.0926 -  
accuracy: 0.9667 - val_loss: 0.3172 - val_accuracy: 0.9056  
Epoch 4/10  
620/620 [=====] - 10s 17ms/step - loss: 0.0476 -  
accuracy: 0.9835 - val_loss: 0.4106 - val_accuracy: 0.8999  
Epoch 5/10  
620/620 [=====] - 10s 17ms/step - loss: 0.0323 -  
accuracy: 0.9890 - val_loss: 0.4569 - val_accuracy: 0.9003  
Epoch 6/10  
620/620 [=====] - 10s 17ms/step - loss: 0.0245 -  
accuracy: 0.9921 - val_loss: 0.4946 - val_accuracy: 0.9040  
Epoch 7/10  
620/620 [=====] - 10s 17ms/step - loss: 0.0223 -  
accuracy: 0.9915 - val_loss: 0.4960 - val_accuracy: 0.8935  
Epoch 8/10  
620/620 [=====] - 10s 17ms/step - loss: 0.0215 -  
accuracy: 0.9921 - val_loss: 0.5403 - val_accuracy: 0.9018  
Epoch 9/10  
620/620 [=====] - 10s 17ms/step - loss: 0.0176 -  
accuracy: 0.9929 - val_loss: 0.6062 - val_accuracy: 0.8731  
Epoch 10/10  
620/620 [=====] - 11s 17ms/step - loss: 0.0151 -  
accuracy: 0.9935 - val_loss: 0.7036 - val_accuracy: 0.8973
```

```
[192]: <keras.src.callbacks.History at 0x178ecceca90>
```

0.3.4 Getting Representation From the Model

```
[203]: cnn_embeddings = model.layers[0].get_weights()[0]
cnn_embeddings
```

```
[203]: array([[ -0.02070224, -0.00217342, -0.01971866, ...,  0.00169    ,
          -0.00793593,  0.00185298],
          [ 0.1213275 , -0.11898055, -0.06785338, ..., -0.08902466,
          -0.08240429,  0.14615555],
          [-0.03989461, -0.0157681 ,  0.04168139, ...,  0.01872528,
          -0.02950534, -0.01898918],
          ...,
          [ 0.02673351,  0.02214083, -0.01881211, ..., -0.04175155,
           0.02514503,  0.03081172],
          [-0.01907891, -0.01436325,  0.02728608, ...,  0.02839203,
          -0.00134127, -0.00377864],
          [-0.01459378,  0.0241019 , -0.03598715, ..., -0.04778757,
          -0.00159583, -0.04600607]], dtype=float32)
```

```
[207]: cnn_embeddings.shape
```

```
[207]: (16841, 100)
```

0.4 Downstream Evaluation

```
[224]: def train_and_evaluate_decision_tree(x_train, x_test, y_train, y_test,
      ↪representation):

    dtclassifier = DecisionTreeClassifier(random_state=9,max_depth=5)
    dtclassifier.fit(x_train, y_train)
    y_pred = dtclassifier.predict(x_test)

    print(f"\nMetrics for {representation}:")
    print(f"Model Score: {dtclassifier.score(x_train,y_train)}")
    print("Accuracy:", accuracy_score(y_test, y_pred))
    print("Classification Report:\n", classification_report(y_test, y_pred))
```

```
[228]: def get_dataset_prep(array,labels):
    x_train,x_test,y_train,y_test = train_test_split(array,labels,test_size=0.
      ↪2,random_state=2)
    return x_train,x_test,y_train,y_test
```

0.4.1 Decision Tree - CBOW

```
[229]: x_train,x_test,y_train,y_test = get_dataset_prep(cbow_array,labels)
```

```
[237]: train_and_evaluate_decision_tree(x_train, x_test, y_train, y_test, "DT-CBOW")
```

Metrics for DT:

Model Score: 0.7830517023959647

Accuracy: 0.7732499495662699

Classification Report:

	precision	recall	f1-score	support
0	0.53	0.03	0.05	279
1	0.78	0.99	0.87	3835
2	0.45	0.05	0.09	843
accuracy			0.77	4957
macro avg	0.59	0.35	0.34	4957
weighted avg	0.71	0.77	0.69	4957

0.4.2 Decision Tree - Skipgram

```
[238]: x_train,x_test,y_train,y_test = get_dataset_prep(sg_array,labels)
```

```
[239]: train_and_evaluate_decision_tree(x_train, x_test, y_train, y_test,□  
      ↪ "DT-SkipGram")
```

Metrics for DT-SkipGram:

Model Score: 0.7810844892812105

Accuracy: 0.7758725035303611

Classification Report:

	precision	recall	f1-score	support
0	0.42	0.03	0.05	279
1	0.79	0.98	0.87	3835
2	0.50	0.12	0.19	843
accuracy			0.78	4957
macro avg	0.57	0.37	0.37	4957
weighted avg	0.72	0.78	0.71	4957

0.5 RNN - Representation

```
[240]: vocab_size = len(tokenizer.word_index) + 1  
      vocab_size
```

```
[240]: 16841
```

```
[241]: rnn_representation_dim = 100
```

```
[244]: model = Sequential()
model.add(Embedding(vocab_size, rnn_representation_dim, input_length=max_len))
model.add(SimpleRNN(32, return_sequences=False))
model.add(Dense(128, activation='relu'))
model.add(Dense(3, activation='softmax'))
```

```
[245]: model.summary()
```

Model: "sequential_7"

Layer (type)	Output Shape	Param #
embedding_4 (Embedding)	(None, 138, 100)	1684100
simple_rnn (SimpleRNN)	(None, 32)	4256
dense_5 (Dense)	(None, 128)	4224
dense_6 (Dense)	(None, 3)	387

Total params: 1692967 (6.46 MB)
 Trainable params: 1692967 (6.46 MB)
 Non-trainable params: 0 (0.00 Byte)

```
[246]: model.compile(optimizer='adam', loss='sparse_categorical_crossentropy',
↪ metrics=['accuracy'])
```

0.5.1 Model Training

```
[247]: model.fit(tweet_data, labels, epochs=10, batch_size=32, validation_split=0.2)
```

```
Epoch 1/10
620/620 [=====] - 14s 21ms/step - loss: 0.4354 -
accuracy: 0.8473 - val_loss: 0.2882 - val_accuracy: 0.9074
Epoch 2/10
620/620 [=====] - 12s 20ms/step - loss: 0.2148 -
accuracy: 0.9238 - val_loss: 0.2949 - val_accuracy: 0.9058
Epoch 3/10
620/620 [=====] - 12s 20ms/step - loss: 0.0923 -
accuracy: 0.9682 - val_loss: 0.3596 - val_accuracy: 0.8963
Epoch 4/10
620/620 [=====] - 12s 20ms/step - loss: 0.0504 -
accuracy: 0.9831 - val_loss: 0.4503 - val_accuracy: 0.8798
Epoch 5/10
620/620 [=====] - 12s 20ms/step - loss: 0.0360 -
```



```

accuracy: 0.9875 - val_loss: 0.4897 - val_accuracy: 0.8800
Epoch 6/10
620/620 [=====] - 12s 20ms/step - loss: 0.0283 -
accuracy: 0.9900 - val_loss: 0.5548 - val_accuracy: 0.8737
Epoch 7/10
620/620 [=====] - 12s 20ms/step - loss: 0.0245 -
accuracy: 0.9909 - val_loss: 0.6313 - val_accuracy: 0.8598
Epoch 8/10
620/620 [=====] - 12s 20ms/step - loss: 0.0219 -
accuracy: 0.9920 - val_loss: 0.6256 - val_accuracy: 0.8646
Epoch 9/10
620/620 [=====] - 12s 20ms/step - loss: 0.0258 -
accuracy: 0.9904 - val_loss: 0.6873 - val_accuracy: 0.8671
Epoch 10/10
620/620 [=====] - 12s 20ms/step - loss: 0.0195 -
accuracy: 0.9927 - val_loss: 0.7399 - val_accuracy: 0.8527

```

[247]: <keras.src.callbacks.History at 0x178f2f7dd90>

0.5.2 Getting Representation from Model

```

[248]: rnn_embeddings = model.layers[0].get_weights()[0]
       rnn_embeddings

```

```

[248]: array([[ 0.08894386,  0.05566104,  0.0040675 , ...,  0.00197502,
                -0.05583239, -0.00615073],
              [-0.0493452 , -0.17989728, -0.08420745, ...,  0.27601382,
                -0.00838685,  0.21541148],
              [-0.01406219, -0.05131751, -0.01463335, ...,  0.03848261,
                0.00777662,  0.00761374],
              ...,
              [-0.01450269,  0.02662939, -0.01659643, ..., -0.02704675,
                0.04928735, -0.00095978],
              [ 0.02304536, -0.00220132,  0.02080033, ..., -0.01050389,
                -0.02484881, -0.00413143],
              [-0.04524095,  0.01009665,  0.04885234, ...,  0.03640784,
                -0.01080083, -0.01088579]], dtype=float32)

```