

# Assignment4

November 1, 2022

## 0.1 ASSIGNMENT 4

## 0.2 Team ID - PNT2022TMID38424

## 0.3 Project - A Novel Method for Handwritten Digit Recognition System

## 0.4 Import required library

```
[33]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import tensorflow
import keras
import re
import nltk
from nltk.corpus import stopwords
from nltk.stem.porter import PorterStemmer
from wordcloud import WordCloud
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import LSTM, Dense, Dropout, Embedding
from tensorflow.keras.optimizers import RMSprop
from tensorflow.keras.preprocessing.text import Tokenizer
from tensorflow.keras.preprocessing import sequence
```

## 1 Read Dataset and do Pre-processing

```
[34]: file = '/content/spam.tsv'
dataset = pd.read_csv(file, sep="\t")
```

```
[35]: dataset.head()
```

```
[35]:      v1                                     v2 Unnamed: 2  \
0  ham  Go until jurong point, crazy.. Available only ...      NaN
```

1	ham	Ok lar... Joking wif u oni...	NaN
2	spam	Free entry in 2 a wkly comp to win FA Cup fina...	NaN
3	ham	U dun say so early hor... U c already then say...	NaN
4	ham	Nah I don't think he goes to usf, he lives aro...	NaN

```

    Unnamed: 3 Unnamed: 4
0      NaN      NaN
1      NaN      NaN
2      NaN      NaN
3      NaN      NaN
4      NaN      NaN

```

```
[36]: dataset.drop(['Unnamed: 2', 'Unnamed: 3', 'Unnamed: 4'], axis=1, inplace=True)
dataset.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5572 entries, 0 to 5571
Data columns (total 2 columns):
#   Column  Non-Null Count  Dtype
---  -
0    v1      5572 non-null     object
1    v2      5572 non-null     object
dtypes: object(2)
memory usage: 87.2+ KB

```

```
[37]: ps = PorterStemmer()
```

```
[38]: review = dataset["v2"][0]
review
```

```
[38]: 'Go until jurong point, crazy.. Available only in bugis n great world la e
buffet... Cine there got amore wat...'
```

```
[39]: review = review.lower()
review
```

```
[39]: 'go until jurong point, crazy.. available only in bugis n great world la e
buffet... cine there got amore wat...'
```

```
[40]: review = review.split()
review
```

```
[40]: ['go',
      'until',
      'jurong',
      'point,',
      'crazy..',
```

```

'available',
'only',
'in',
'bugis',
'n',
'great',
'world',
'la',
'e',
'buffet...',
'cine',
'there',
'got',
'amore',
'wat...']

```

```

[42]: nltk.download('stopwords')
review = [ps.stem(word) for word in review if not word in set(stopwords.
↳ words('english'))]
review

```

```

[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Unzipping corpora/stopwords.zip.

```

```

[42]: ['go',
'jurong',
'point,',
'crazy..',
'avail',
'bugi',
'n',
'great',
'world',
'la',
'e',
'buffet...',
'cine',
'got',
'amor',
'wat...']

```

```

[43]: sw = stopwords.words("english")
'wat' in sw

```

```

[43]: False

```

```

[44]: dataset['v1'].value_counts()

```

```
[44]: ham      4825  
      spam      747  
      Name: v1, dtype: int64
```

```
[45]: corpus = " ".join(dataset['v2'])
```

```
[46]: corpus.split()
```

```
[46]: ['Go',  
      'until',  
      'jurong',  
      'point',  
      'crazy..',  
      'Available',  
      'only',  
      'in',  
      'bugis',  
      'n',  
      'great',  
      'world',  
      'la',  
      'e',  
      'buffet...',  
      'Cine',  
      'there',  
      'got',  
      'amore',  
      'wat...',  
      'Ok',  
      'lar...',  
      'Joking',  
      'wif',  
      'u',  
      'oni...',  
      'Free',  
      'entry',  
      'in',  
      '2',  
      'a',  
      'wkly',  
      'comp',  
      'to',  
      'win',  
      'FA',  
      'Cup',  
      'final',  
      'tkts',
```

'21st',  
'May',  
'2005.',  
'Text',  
'FA',  
'to',  
'87121',  
'to',  
'receive',  
'entry',  
'question(std',  
'txt',  
'rate)T&C's",  
'apply',  
'08452810075over18's",  
'U',  
'dun',  
'say',  
'so',  
'early',  
'hor...',  
'U',  
'c',  
'already',  
'then',  
'say...',  
'Nah',  
'I',  
'don't",  
'think',  
'he',  
'goes',  
'to',  
'usf,',  
'he',  
'lives',  
'around',  
'here',  
'though',  
'FreeMsg',  
'Hey',  
'there',  
'darling',  
'it's",  
'been',  
'3',  
'week's",

'now',  
'and',  
'no',  
'word',  
'back!',  
"I'd",  
'like',  
'some',  
'fun',  
'you',  
'up',  
'for',  
'it',  
'still?',  
'Tb',  
'ok!',  
'XxX',  
'std',  
'chgs',  
'to',  
'send,',  
' 1.50',  
'to',  
'rcv',  
'Even',  
'my',  
'brother',  
'is',  
'not',  
'like',  
'to',  
'speak',  
'with',  
'me.',  
'They',  
'treat',  
'me',  
'like',  
'aids',  
'patent.',  
'As',  
'per',  
'your',  
'request',  
"Melle",  
'Melle',  
'(Oru',

'Minnaminunginte',  
'Nurungu',  
"Vettam)"',  
'has',  
'been',  
'set',  
'as',  
'your',  
'callertune',  
'for',  
'all',  
'Callers.',  
'Press',  
'\*9',  
'to',  
'copy',  
'your',  
'friends',  
'Callertune',  
'WINNER!!',  
'As',  
'a',  
'valued',  
'network',  
'customer',  
'you',  
'have',  
'been',  
'selected',  
'to',  
'receivea',  
' 900',  
'prize',  
'reward!',  
'To',  
'claim',  
'call',  
'09061701461.',  
'Claim',  
'code',  
'KL341.',  
'Valid',  
'12',  
'hours',  
'only.',  
'Had',  
'your',

'mobile',  
'11',  
'months',  
'or',  
'more?',  
'U',  
'R',  
'entitled',  
'to',  
'Update',  
'to',  
'the',  
'latest',  
'colour',  
'mobiles',  
'with',  
'camera',  
'for',  
'Free!',  
'Call',  
'The',  
'Mobile',  
'Update',  
'Co',  
'FREE',  
'on',  
'08002986030',  
"I'm",  
'gonna',  
'be',  
'home',  
'soon',  
'and',  
'i',  
"don't",  
'want',  
'to',  
'talk',  
'about',  
'this',  
'stuff',  
'anymore',  
'tonight',  
'k?',  
"I've",  
'cried',  
'enough',



'today.',  
'SIX',  
'chances',  
'to',  
'win',  
'CASH!',  
'From',  
'100',  
'to',  
'20,000',  
'pounds',  
'txt>',  
'CSH11',  
'and',  
'send',  
'to',  
'87575.',  
'Cost',  
'150p/day,',  
'6days,',  
'16+',  
'TsandCs',  
'apply',  
'Reply',  
'HL',  
'4',  
'info',  
'URGENT!',  
'You',  
'have',  
'won',  
'a',  
'1',  
'week',  
'FREE',  
'membership',  
'in',  
'our',  
' 100,000',  
'Prize',  
'Jackpot!',  
'Txt',  
'the',  
'word:',  
'CLAIM',  
'to',  
'No:',

'81010',  
'T&C',  
'www.dbuk.net',  
'LCCLTD',  
'POBOX',  
'4403LDNW1A7RW18',  
"I've",  
'been',  
'searching',  
'for',  
'the',  
'right',  
'words',  
'to',  
'thank',  
'you',  
'for',  
'this',  
'breather.',  
'I',  
'promise',  
'i',  
'wont',  
'take',  
'your',  
'help',  
'for',  
'granted',  
'and',  
'will',  
'fulfil',  
'my',  
'promise.',  
'You',  
'have',  
'been',  
'wonderful',  
'and',  
'a',  
'blessing',  
'at',  
'all',  
'times.',  
'I',  
'HAVE',  
'A',  
'DATE',

'ON',  
'SUNDAY',  
'WITH',  
'WILL!!',  
'XXXMobileMovieClub:',  
'To',  
'use',  
'your',  
'credit',  
'click',  
'the',  
'WAP',  
'link',  
'in',  
'the',  
'next',  
'txt',  
'message',  
'or',  
'click',  
'here>>',  
'http://wap.',  
'xxmobilemovieclub.com?n=QJKGIGHJJGCBL',  
'Oh',  
'k...i'm",  
'watching',  
'here:)',  
'Eh',  
'u',  
'remember',  
'how',  
'2',  
'spell',  
'his',  
'name...',  
'Yes',  
'i',  
'did.',  
'He',  
'v',  
'naughty',  
'make',  
'until',  
'i',  
'v',  
'wet.',  
'Fine',

'if',  
'that s',  
'the',  
'way',  
'u',  
'feel.',  
'That s',  
'the',  
'way',  
'its',  
'gota',  
'b',  
'England',  
'v',  
'Macedonia',  
'-',  
'dont',  
'miss',  
'the',  
'goals/team',  
'news.',  
'Txt',  
'ur',  
'national',  
'team',  
'to',  
'87077',  
'eg',  
'ENGLAND',  
'to',  
'87077',  
'Try:WALES,',  
'SCOTLAND',  
'4txt/1.20',  
'POBOXox36504W45WQ',  
'16+',  
'Is',  
'that',  
'seriously',  
'how',  
'you',  
'spell',  
'his',  
'name?',  
'I m',  
'going',  
'to',

'try',  
'for',  
'2',  
'months',  
'ha',  
'ha',  
'only',  
'joking',  
'So',  
'-',  
'pay',  
'first',  
'lar...',  
'Then',  
'when',  
'is',  
'da',  
'stock',  
'comin...',  
'Aft',  
'i',  
'finish',  
'my',  
'lunch',  
'then',  
'i',  
'go',  
'str',  
'down',  
'lor.',  
'Ard',  
'3',  
'smth',  
'lor.',  
'U',  
'finish',  
'ur',  
'lunch',  
'already?',  
'Ffffffff.',  
'Alright',  
'no',  
'way',  
'I',  
'can',  
'meet',  
'up',

'with',  
'you',  
'sooner?',  
'Just',  
'forced',  
'myself',  
'to',  
'eat',  
'a',  
'slice.',  
'I'm',  
'really',  
'not',  
'hungry',  
'tho.',  
'This',  
'sucks.',  
'Mark',  
'is',  
'getting',  
'worried.',  
'He',  
'knows',  
'I'm',  
'sick',  
'when',  
'I',  
'turn',  
'down',  
'pizza.',  
'Lol',  
'Lol',  
'your',  
'always',  
'so',  
'convincing.',  
'Did',  
'you',  
'catch',  
'the',  
'bus',  
'?',  
'Are',  
'you',  
'frying',  
'an',  
'egg',

'?',  
'Did',  
'you',  
'make',  
'a',  
'tea?',  
'Are',  
'you',  
'eating',  
'your',  
"mom's",  
'left',  
'over',  
'dinner',  
'?',  
'Do',  
'you',  
'feel',  
'my',  
'Love',  
'?',  
"I'm",  
'back',  
'&',  
"we're",  
'packing',  
'the',  
'car',  
'now',  
"I'll",  
'let',  
'you',  
'know',  
'if',  
"there's",  
'room',  
'Ahhh.',  
'Work.',  
'I',  
'vaguely',  
'remember',  
'that!',  
'What',  
'does',  
'it',  
'feel',  
'like?'

'Lol',  
'Wait',  
"that's",  
'still',  
'not',  
'all',  
'that',  
'clear',  
'were',  
'you',  
'not',  
'sure',  
'about',  
'me',  
'being',  
'sarcastic',  
'or',  
'that',  
"that's",  
'why',  
'x',  
"doesn't",  
'want',  
'to',  
'live',  
'with',  
'us',  
'Yeah',  
'he',  
'got',  
'in',  
'at',  
'2',  
'and',  
'was',  
'v',  
'apologetic.',  
'n',  
'had',  
'fallen',  
'out',  
'and',  
'she',  
'was',  
'actin',  
'like',  
'spoilt',



'child',  
'and',  
'he',  
'got',  
'caught',  
'up',  
'in',  
'that.',  
'Till',  
'2!',  
'But',  
'we',  
'won't',  
'go',  
'there!',  
'Not',  
'doing',  
'too',  
'badly',  
'cheers.',  
'You?',  
'K',  
'tell',  
'me',  
'anything',  
'about',  
'you.',  
'For',  
'fear',  
'of',  
'fainting',  
'with',  
'the',  
'of',  
'all',  
'that',  
'housework',  
'you',  
'just',  
'did?',  
'Quick',  
'have',  
'a',  
'cuppa',  
'Thanks',  
'for',  
'your',

'subscription',  
'to',  
'Ringtone',  
'UK',  
'your',  
'mobile',  
'will',  
'be',  
'charged',  
' 5/month',  
'Please',  
'confirm',  
'by',  
'replying',  
'YES',  
'or',  
'NO.',  
'If',  
'you',  
'reply',  
'NO',  
'you',  
'will',  
'not',  
'be',  
'charged',  
'Yup...',  
'Ok',  
'i',  
'go',  
'home',  
'look',  
'at',  
'the',  
'timings',  
'then',  
'i',  
'msg',  
'\_',  
'again...',  
'Xuhui',  
'going',  
'to',  
'learn',  
'on',  
'2nd',  
'may',

'too',  
'but',  
'her',  
'lesson',  
'is',  
'at',  
'8am',  
'Oops',  
"I'll",  
'let',  
'you',  
'know',  
'when',  
'my',  
"roommate's",  
'done',  
'I',  
'see',  
'the',  
'letter',  
'B',  
'on',  
'my',  
'car',  
'Anything',  
'lor...',  
'U',  
'decide...',  
'Hello!',  
"How's",  
'you',  
'and',  
'how',  
'did',  
'saturday',  
'go?',  
'I',  
'was',  
'just',  
'texting',  
'to',  
'see',  
'if',  
"you'd",  
'decided',  
'to',  
'do',

'anything',  
'tomo.',  
'Not',  
'that',  
"i'm",  
'trying',  
'to',  
'invite',  
'myself',  
'or',  
'anything!'  
'Pls',  
'go',  
'ahead',  
'with',  
'watts.',  
'I',  
'just',  
'wanted',  
'to',  
'be',  
'sure.',  
'Do',  
'have',  
'a',  
'great',  
'weekend.',  
'Abiola',  
'Did',  
'I',  
'forget',  
'to',  
'tell',  
'you',  
'?',  
'I',  
'want',  
'you',  
'',  
'I',  
'need',  
'you',  
'I',  
'crave',  
'you',  
'...'  
'But',

'most',  
'of',  
'all',  
'...',  
'I',  
'love',  
'you',  
'my',  
'sweet',  
'Arabian',  
'steed',  
'...',  
'Mmmmmmm',  
'...',  
'Yummy',  
'07732584351',  
'-',  
'Rodger',  
'Burns',  
'-',  
'MSG',  
'=',  
'We',  
'tried',  
'to',  
'call',  
'you',  
're',  
'your',  
'reply',  
'to',  
'our',  
'sms',  
'for',  
'a',  
'free',  
'nokia',  
'mobile',  
'+',  
'free',  
'camcorder.',  
'Please',  
'call',  
'now',  
'08000930705',  
'for',  
'delivery',

'tomorrow',  
'WHO',  
'ARE',  
'YOU',  
'SEEING?',  
'Great!',  
'I',  
'hope',  
'you',  
'like',  
'your',  
'man',  
'well',  
'endowed.',  
'I',  
'am',  
'&lt;#&gt;',  
'inches...',  
'No',  
'calls..messages..missed',  
'calls',  
'Didn't',  
'you',  
'get',  
'hep',  
'b',  
'immunisation',  
'in',  
'nigeria.',  
'Fair',  
'enough',  
'anything',  
'going',  
'on?',  
'Yeah',  
'hopefully',  
'if',  
'tyler',  
'can't',  
'do',  
'it',  
'I',  
'could',  
'maybe',  
'ask',  
'around',  
'a',

'bit',  
'U',  
"don't",  
'know',  
'how',  
'stubborn',  
'I',  
'am.',  
'I',  
"didn't",  
'even',  
'want',  
'to',  
'go',  
'to',  
'the',  
'hospital.',  
'I',  
'kept',  
'telling',  
'Mark',  
"I'm",  
'not',  
'a',  
'weak',  
'sucker.',  
'Hospitals',  
'are',  
'for',  
'weak',  
'suckers.',  
'What',  
'you',  
'thought',  
'about',  
'me.',  
'First',  
'time',  
'you',  
'saw',  
'me',  
'in',  
'class.',  
'A',  
'gram',  
'usually',  
'runs',

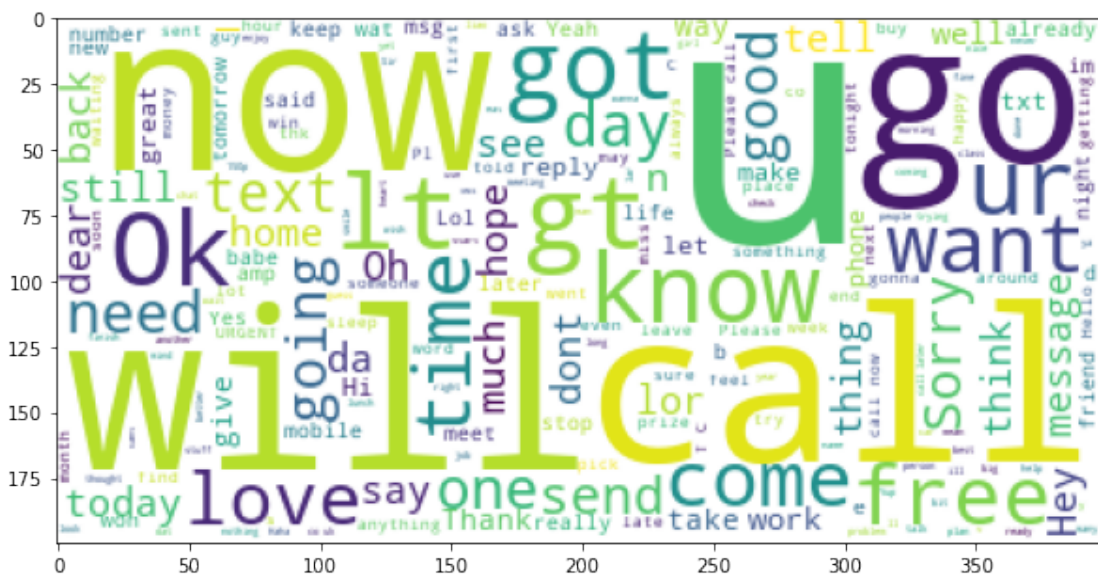
'like',  
'&lt;#&gt;','  
,',  
'a',  
'half',  
'eighth',  
'is',  
'smarter',  
'though',  
'and',  
'gets',  
'you',  
'almost',  
'a',  
'whole',  
'second',  
'gram',  
'for',  
'&lt;#&gt;','  
'K',  
'fyi',  
'x',  
'has',  
'a',  
'ride',  
'early',  
'tomorrow',  
'morning',  
'but',  
'he's',  
'crashing',  
'at',  
'our',  
'place',  
'tonight',  
'Wow.',  
'I',  
'never',  
'realized',  
'that',  
'you',  
'were',  
'so',  
'embarrassed',  
'by',  
'your',  
'accomodations.',



I',  
'thought',  
'you',  
'liked',  
'it,',  
'since',  
'i',  
'was',  
'doing',  
'the',  
'best',  
'i',  
'could',  
'and',  
'you',  
'always',  
'seemed',  
'so',  
'happy',  
'about',  
'\\the',  
...]

```
[47]: wc = WordCloud( background_color='white')
      review_wordcloud = wc.generate(corpus)
```

```
[50]: plt.figure(figsize=(10,10))
      plt.imshow(review_wordcloud)
      plt.show()
```



```
[51]: word_count = {}  
      for word in corpus.split():  
          if word in word_count:  
              word_count[word] = word_count[word] + 1  
          else:  
              word_count[word] = 1
```

```
[52]: word_count
```

```
[52]: {'Go': 14,  
      'until': 26,  
      'jurong': 1,  
      'point.': 1,  
      'crazy..': 1,  
      'Available': 3,  
      'only': 147,  
      'in': 798,  
      'bugis': 4,  
      'n': 137,  
      'great': 75,  
      'world': 17,  
      'la': 2,  
      'e': 71,  
      'buffet...': 1,  
      'Cine': 1,  
      'there': 118,  
      'got': 204,  
      'amore': 1,  
      'wat...': 12,  
      'Ok': 97,  
      'lar...': 18,  
      'Joking': 1,  
      'wif': 26,  
      'u': 692,  
      'oni...': 2,  
      'Free': 39,  
      'entry': 25,  
      '2': 453,  
      'a': 1327,  
      'wkly': 10,  
      'comp': 10,  
      'to': 2134,  
      'win': 38,  
      'FA': 4,  
      'Cup': 3,
```

'final': 13,  
'tkts': 4,  
'21st': 3,  
'May': 16,  
'2005.': 3,  
'Text': 42,  
'87121': 4,  
'receive': 34,  
'question(std': 2,  
'txt': 75,  
'rate)T&C's": 2,  
'apply': 15,  
'08452810075over18's": 2,  
'U': 299,  
'dun': 46,  
'say': 72,  
'so': 294,  
'early': 24,  
'hor...': 2,  
'c': 44,  
'already': 45,  
'then': 146,  
'say...': 3,  
'Nah': 7,  
'I': 1466,  
'don't": 110,  
'think': 113,  
'he': 127,  
'goes': 26,  
'usf,': 2,  
'lives': 2,  
'around': 53,  
'here': 75,  
'though': 15,  
'FreeMsg': 5,  
'Hey': 76,  
'darling': 4,  
'it's": 53,  
'been': 120,  
'3': 51,  
'week's": 4,  
'now': 190,  
'and': 857,  
'no': 124,  
'word': 22,  
'back!': 2,  
'I'd": 6,

'like': 222,  
'some': 114,  
'fun': 22,  
'you': 1622,  
'up': 241,  
'for': 644,  
'it': 396,  
'still?': 2,  
'Tb': 1,  
'ok!': 2,  
'XxX': 2,  
'std': 2,  
'chgs': 1,  
'send,': 2,  
' 1.50': 17,  
'rcv': 2,  
'Even': 15,  
'my': 628,  
'brother': 9,  
'is': 781,  
'not': 340,  
'speak': 17,  
'with': 349,  
'me.': 97,  
'They': 32,  
'treat': 11,  
'me': 557,  
'aids': 1,  
'patent.': 1,  
'As': 39,  
'per': 55,  
'your': 560,  
'request': 7,  
"'Melle": 3,  
'Melle': 3,  
'(Oru': 3,  
'Minnaminunginte': 3,  
'Nurungu': 3,  
"Vettam)": 3,  
'has': 113,  
'set': 19,  
'as': 133,  
'callertune': 5,  
'all': 216,  
'Callers.': 5,  
'Press': 8,  
'\*9': 5,

'copy': 9,  
'friends': 29,  
'Callertune': 5,  
'WINNER!!': 2,  
'valued': 11,  
'network': 23,  
'customer': 34,  
'have': 474,  
'selected': 27,  
'receivea': 2,  
'900': 7,  
'prize': 44,  
'reward!': 3,  
'To': 79,  
'claim': 73,  
'call': 376,  
'09061701461.': 2,  
'Claim': 30,  
'code': 12,  
'KL341.': 2,  
'Valid': 20,  
'12': 8,  
'hours': 17,  
'only.': 15,  
'Had': 20,  
'mobile': 94,  
'11': 7,  
'months': 7,  
'or': 373,  
'more?': 3,  
'R': 40,  
'entitled': 8,  
'Update': 13,  
'the': 1197,  
'latest': 27,  
'colour': 19,  
'mobiles': 7,  
'camera': 20,  
'Free!': 6,  
'Call': 155,  
'The': 95,  
'Mobile': 30,  
'Co': 5,  
'FREE': 90,  
'on': 487,  
'08002986030': 2,  
'I'm": 286,

'gonna': 57,  
'be': 362,  
'home': 107,  
'soon': 28,  
'i': 742,  
'want': 157,  
'talk': 35,  
'about': 150,  
'this': 223,  
'stuff': 32,  
'anymore': 2,  
'tonight,': 3,  
'k?': 2,  
"I've": 41,  
'cried': 1,  
'enough': 22,  
'today.': 27,  
'SIX': 2,  
'chances': 2,  
'CASH!': 2,  
'From': 13,  
'100': 16,  
'20,000': 2,  
'pounds': 19,  
'txt>': 2,  
'CSH11': 2,  
'send': 150,  
'87575.': 4,  
'Cost': 15,  
'150p/day,': 2,  
'6days,': 2,  
'16+': 22,  
'TsandCs': 2,  
'Reply': 51,  
'HL': 6,  
'4': 249,  
'info': 8,  
'URGENT!': 31,  
'You': 263,  
'won': 45,  
'1': 64,  
'week': 68,  
'membership': 3,  
'our': 126,  
' 100,000': 2,  
'Prize': 14,  
'Jackpot!': 2,

'Txt': 67,  
'word:': 9,  
'CLAIM': 3,  
'No:': 17,  
'81010': 2,  
'T&C': 12,  
'www.dbuk.net': 2,  
'LCCLTD': 2,  
'POBOX': 6,  
'4403LDNW1A7RW18': 2,  
'searching': 5,  
'right': 60,  
'words': 14,  
'thank': 9,  
'breather.': 1,  
'promise': 7,  
'wont': 29,  
'take': 101,  
'help': 29,  
'granted': 1,  
'will': 330,  
'fulfil': 1,  
'promise.': 2,  
'wonderful': 13,  
'blessing': 2,  
'at': 359,  
'times.': 4,  
'HAVE': 11,  
'A': 92,  
'DATE': 5,  
'ON': 18,  
'SUNDAY': 1,  
'WITH': 14,  
'WILL!!': 1,  
'XXXMobileMovieClub.': 1,  
'use': 37,  
'credit,': 1,  
'click': 5,  
'WAP': 6,  
'link': 8,  
'next': 57,  
'message': 56,  
'here>>': 1,  
'http://wap.': 1,  
'xxmobilemovieclub.com?n=QJKGIGHJJGCBL': 1,  
'Oh': 61,  
"k...i'm": 1,

'watching': 22,  
'here:)': 2,  
'Eh': 6,  
'remember': 23,  
'how': 111,  
'spell': 3,  
'his': 49,  
'name...': 1,  
'Yes': 36,  
'did.': 2,  
'He': 53,  
'v': 28,  
'naughty': 5,  
'make': 90,  
'wet.': 1,  
'Fine': 3,  
'if': 235,  
'that s': 1,  
'way': 75,  
'feel.': 1,  
'That s': 1,  
'its': 128,  
'gota': 1,  
'b': 51,  
'England': 4,  
'Macedonia': 1,  
'-': 88,  
'dont': 97,  
'miss': 60,  
'goals/team': 1,  
'news.': 1,  
'ur': 293,  
'national': 7,  
'team': 3,  
'87077': 7,  
'eg': 11,  
'ENGLAND': 3,  
'Try:WALES,': 1,  
'SCOTLAND': 1,  
'4txt/1.20': 1,  
'POBOXox36504W45WQ': 1,  
'Is': 52,  
'that': 421,  
'seriously': 4,  
'name?': 1,  
'I m': 4,  
'going': 142,



'try': 36,  
'ha': 7,  
'joking': 2,  
'So': 118,  
'\_': 105,  
'pay': 28,  
'first': 41,  
'Then': 57,  
'when': 233,  
'da': 72,  
'stock': 5,  
'comin...': 1,  
'Aft': 4,  
'finish': 34,  
'lunch': 32,  
'go': 234,  
'str': 3,  
'down': 45,  
'lor.': 67,  
'Ard': 9,  
'smth': 12,  
'already?': 14,  
'Ffffffff.': 1,  
'Alright': 7,  
'can': 304,  
'meet': 71,  
'sooner?': 1,  
'Just': 119,  
'forced': 1,  
'myself': 8,  
'eat': 30,  
'slice.': 1,  
'really': 69,  
'hungry': 8,  
'tho.': 3,  
'This': 77,  
'sucks.': 4,  
'Mark': 5,  
'getting': 46,  
'worried.': 3,  
'knows': 8,  
'sick': 7,  
'turn': 3,  
'pizza.': 2,  
'Lol': 48,  
'always': 45,  
'convincing.': 1,

'Did': 51,  
'catch': 8,  
'bus': 21,  
'?': 187,  
'Are': 78,  
'frying': 3,  
'an': 108,  
'egg': 2,  
'tea?': 1,  
'eating': 5,  
'mom's': 4,  
'left': 23,  
'over': 43,  
'dinner': 24,  
'Do': 113,  
'feel': 55,  
'Love': 25,  
'back': 111,  
'&': 82,  
'we're': 16,  
'packing': 1,  
'car': 33,  
'now,': 16,  
'I'll': 137,  
'let': 54,  
'know': 224,  
'there's': 16,  
'room': 24,  
'Ahhh.': 1,  
'Work.': 1,  
'vaguely': 1,  
'that!': 7,  
'What': 101,  
'does': 20,  
'like?': 4,  
'Wait': 6,  
'that's': 30,  
'still': 134,  
'clear,': 1,  
'were': 53,  
'sure': 53,  
'being': 42,  
'sarcastic': 2,  
'why': 46,  
'x': 33,  
'doesn't': 12,  
'live': 27,

'us': 51,  
'Yeah': 43,  
'was': 200,  
'apologetic.': 1,  
'had': 76,  
'fallen': 1,  
'out': 207,  
'she': 85,  
'actin': 1,  
'spoilt': 1,  
'child': 3,  
'caught': 2,  
'that.': 23,  
'Till': 1,  
'2!': 3,  
'But': 131,  
'we': 221,  
"won't": 22,  
'there!': 6,  
'Not': 56,  
'doing': 64,  
'too': 69,  
'badly': 1,  
'cheers.': 1,  
'You?': 1,  
'K': 28,  
'tell': 104,  
'anything': 48,  
'you.': 82,  
'For': 40,  
'fear': 1,  
'of': 591,  
'fainting': 1,  
'housework': 1,  
'just': 236,  
'did?': 1,  
'Quick': 1,  
'cuppa': 1,  
'Thanks': 41,  
'subscription': 3,  
'Ringtone': 4,  
'UK': 8,  
'charged': 11,  
'5/month': 1,  
'Please': 64,  
'confirm': 8,  
'by': 155,

'replying': 7,  
'YES': 11,  
'NO.': 2,  
'If': 133,  
'reply': 75,  
'NO': 7,  
'Yup...': 6,  
'look': 27,  
'timings': 1,  
'msg': 49,  
'again...': 6,  
'Xuhui': 2,  
'learn': 7,  
'2nd': 30,  
'may': 27,  
'but': 282,  
'her': 101,  
'lesson': 13,  
'8am': 1,  
'Oops,': 1,  
'roommate's': 2,  
'done': 24,  
'see': 113,  
'letter': 3,  
'B': 12,  
'Anything': 11,  
'lor...': 56,  
'decide...': 1,  
'Hello!': 5,  
'How's': 14,  
'did': 61,  
'saturday': 7,  
'go?': 3,  
'texting': 5,  
'you'd': 6,  
'decided': 15,  
'do': 243,  
'tomo.': 2,  
'i'm': 91,  
'trying': 42,  
'invite': 3,  
'anything!': 1,  
'Pls': 62,  
'ahead': 6,  
'watts.': 1,  
'wanted': 26,  
'sure.': 6,

'weekend.': 8,  
'Abiola': 9,  
'forget': 18,  
' ': 30,  
'need': 141,  
'you,': 33,  
'crave': 11,  
'...': 163,  
'most': 24,  
'love': 114,  
'sweet': 20,  
'Arabian': 1,  
'steed': 1,  
'Mmmmmmm': 2,  
'Yummy': 1,  
'07732584351': 1,  
'Rodger': 1,  
'Burns': 1,  
'MSG': 3,  
'=': 10,  
'We': 119,  
'tried': 22,  
're': 17,  
'sms': 14,  
'free': 99,  
'nokia': 6,  
'+' : 29,  
'camcorder.': 1,  
'08000930705': 14,  
'delivery': 14,  
'tomorrow': 53,  
'WHO': 5,  
'ARE': 11,  
'YOU': 32,  
'SEEING?': 1,  
'Great!': 5,  
'hope': 52,  
'man': 20,  
'well': 39,  
'endowed.': 1,  
'am': 186,  
'&lt;#&gt;': 276,  
'inches...': 1,  
'No': 126,  
'calls..messages..missed': 3,  
'calls': 14,  
"Didn't": 2,

'get': 326,  
'hep': 1,  
'immunisation': 1,  
'nigeria.': 2,  
'Fair': 1,  
'enough.': 1,  
'on?': 2,  
'hopefully.': 1,  
'tyler': 4,  
"can't": 51,  
'could': 56,  
'maybe': 18,  
'ask': 82,  
'bit': 34,  
'stubborn': 1,  
'am.': 3,  
"didn't": 39,  
'even': 45,  
'hospital.': 5,  
'kept': 7,  
'telling': 14,  
'weak': 5,  
'sucker.': 1,  
'Hospitals': 1,  
'are': 396,  
'suckers.': 1,  
'thought': 1,  
'First': 7,  
'time': 154,  
'saw': 21,  
'class.': 14,  
'gram': 3,  
'usually': 4,  
'runs': 3,  
'half': 30,  
'eighth': 1,  
'smarter': 1,  
'gets': 11,  
'almost': 12,  
'whole': 14,  
'second': 13,  
'fyi': 4,  
'ride': 4,  
'morning': 34,  
"he's": 24,  
'crashing': 1,  
'place': 45,

'tonight': 26,  
'Wow.': 1,  
'never': 32,  
'realized': 3,  
'embarrassed': 2,  
'accomodations.': 1,  
'thought': 37,  
'liked': 7,  
'it.': 24,  
'since': 22,  
'best': 31,  
'seemed': 2,  
'happy': 36,  
'\\the': 1,  
'cave\\".': 1,  
'sorry': 31,  
'more': 87,  
'give.': 2,  
'offered.': 1,  
'embarassing."': 1,  
'SMS.': 5,  
'ac': 4,  
'Sptv': 1,  
'New': 29,  
'Jersey': 1,  
'Devils': 1,  
'Detroit': 1,  
'Red': 2,  
'Wings': 1,  
'play': 23,  
'Ice': 1,  
'Hockey.': 1,  
'Correct': 1,  
'Incorrect?': 1,  
'End?': 2,  
'END': 4,  
'SPTV': 1,  
'what': 142,  
'Mallika': 1,  
'Sherawat': 1,  
'yesterday?': 2,  
'Find': 4,  
'@': 11,  
'&lt;URL&gt;': 3,  
'Congrats!': 8,  
'year': 24,  
'special': 29,

'cinema': 5,  
'pass': 9,  
'yours.': 11,  
'09061209465': 3,  
'now!': 53,  
'C': 30,  
'Suprman': 3,  
'V,': 3,  
'Matrix3,': 3,  
'StarWars3,': 3,  
'etc': 7,  
'FREE!': 13,  
'bx420-ip4-5we.': 3,  
'150pm.': 3,  
'Dont': 36,  
'out!': 10,  
'Sorry,': 52,  
'later': 87,  
'meeting.': 3,  
'Tell': 27,  
'where': 56,  
'reached': 9,  
'Yes..gauti': 1,  
'sehwag': 1,  
'odi': 1,  
'series.': 1,  
'Your': 104,  
'pick': 64,  
'\$1': 1,  
'burger': 1,  
'yourself': 14,  
'home.': 30,  
'move.': 1,  
'Pain': 3,  
'killing': 2,  
'Ha': 4,  
'good': 114,  
'joke.': 3,  
'Girls': 1,  
'situation': 5,  
'seekers.': 1,  
'Its': 70,  
'part': 25,  
'checking': 11,  
'IQ': 2,  
'Sorry': 52,  
'roommates': 3,



'took': 17,  
'forever,': 1,  
'ok': 60,  
'come': 198,  
'now?': 25,  
'lar': 10,  
'double': 3,  
'check': 23,  
'hair': 19,  
'dresser': 3,  
'said': 65,  
'wun': 6,  
'cut': 10,  
'short.': 2,  
'nice.': 7,  
'customer,': 5,  
'pleased': 6,  
'advise': 3,  
'following': 7,  
'recent': 3,  
'review': 5,  
'Mob': 3,  
'No.': 22,  
'awarded': 37,  
' 1500': 3,  
'Bonus': 15,  
'Prize,': 4,  
'09066364589': 2,  
'Today': 20,  
'\\song': 2,  
'dedicated': 2,  
'day..\\": 2,  
'Which': 12,  
'song': 9,  
'dedicate': 2,  
'me?': 20,  
'Send': 35,  
'valuable': 3,  
'frnds': 9,  
'rply': 10,  
'me...": 2,  
'Urgent': 6,  
'UR': 29,  
'complimentary': 12,  
'trip': 16,  
'EuroDisinc': 2,  
'Trav,': 2,

'Aco&Entry41': 2,  
'Or': 37,  
' 1000.': 2,  
'DIS': 2,  
'18+6\* 1.50(moreFrmMob.': 2,  
'ShrAcomOrSglSuplt)10,': 2,  
'LS1': 2,  
'3AJ': 2,  
'hear': 26,  
'new': 89,  
'\\Divorce': 1,  
'Barbie\\"?': 1,  
'It': 60,  
'comes': 16,  
'Ken's': 1,  
'stuff!": 1,  
'plane': 3,  
'give': 87,  
'month': 18,  
'end.': 4,  
'Wah': 1,  
'lucky': 11,  
'man...': 1,  
'save': 6,  
'money...': 1,  
'Hee...': 9,  
'Finished': 1,  
'class': 27,  
'HI': 11,  
'BABE': 8,  
'IM': 20,  
'AT': 22,  
'HOME': 3,  
'NOW': 25,  
'WANNA': 3,  
'DO': 9,  
'SOMETHING?': 1,  
'XX': 3,  
'K..k:)where': 1,  
'you?how': 1,  
'performed?': 1,  
'now...': 10,  
'waiting': 43,  
'machan.': 3,  
'once': 22,  
'free.': 6,  
'Thats': 19,

'cool.': 9,  
'gentleman': 3,  
'dignity': 3,  
'respect.': 3,  
'peoples': 1,  
'very': 68,  
'much:)': 1,  
'shy': 1,  
'pa.': 13,  
'Does': 9,  
'operate': 1,  
'after': 70,  
'same': 31,  
'here.': 12,  
'Still': 15,  
'looking': 28,  
'job.': 6,  
'How': 137,  
'much': 83,  
'Ta's': 1,  
'earn': 2,  
'there.': 32,  
'K.': 4,  
'ah?': 14,  
'hi': 32,  
'Yup': 31,  
'stop.': 9,  
'later.': 2,  
'network.': 3,  
'urgnt.': 2,  
'real': 28,  
'yo?': 1,  
'tickets': 7,  
'one': 132,  
'jacket': 2,  
'done.': 6,  
'used': 12,  
'multis.': 1,  
'started': 17,  
'requests': 1,  
'pain': 16,  
'came': 26,  
'bed.': 6,  
'Double': 15,  
'coins': 2,  
'factory': 1,  
'too.': 16,

'gotta': 10,  
'cash': 51,  
'nitros.': 2,  
'babe': 27,  
'Ela': 1,  
'kano.,il': 1,  
'download.': 1,  
'wen': 11,  
'free..': 1,  
'do!': 1,  
'Don t': 2,  
'stand': 7,  
'close': 12,  
'tho-': 1,  
'you ll': 1,  
'something!': 2,  
'pain.': 5,  
'another': 32,  
'night?': 10,  
'spent': 8,  
'late': 39,  
'afternoon': 11,  
'casualty': 1,  
'means': 23,  
'haven't': 17,  
'any': 111,  
'y': 21,  
'stuff42moro': 1,  
'includes': 1,  
'sheets': 2,  
'Sorry.': 7,  
'Smile': 11,  
'Pleasure': 3,  
'trouble': 7,  
'pours': 2,  
'Rain': 2,  
'sum1': 3,  
'Hurts': 2,  
'becoz': 3,  
'SOMEONE': 3,  
'Loves': 5,  
'Smiling!!': 2,  
'service': 45,  
'representative': 10,  
'0800': 13,  
'169': 2,  
'6031': 2,

'between': 31,  
'10am-9pm': 2,  
'WON': 19,  
'guaranteed': 25,  
'1000': 33,  
'5000': 19,  
'prize!': 13,  
'Havent': 5,  
'planning': 8,  
'buy': 54,  
'later.': 24,  
'lido': 1,  
'530': 5,  
'show': 19,  
'afternoon.': 3,  
'work': 62,  
'ringtone': 17,  
'collected.': 2,  
'Simply': 3,  
'text': 126,  
'password': 7,  
'\\MIX\\': 2,  
'85069': 2,  
'verify.': 2,  
'Get': 41,  
'Usher': 2,  
'Britney.': 2,  
'FML': 2,  
'Watching': 12,  
'telugu': 1,  
'movie..wat': 1,  
'abt': 24,  
'u?': 19,  
'see.': 5,  
'When': 46,  
'loads': 10,  
'loans': 1,  
'Hi.': 8,  
'Wk': 2,  
'hols': 2,  
'run.': 2,  
'Forgot': 4,  
'hairstylists': 1,  
'appointment': 3,  
'four': 2,  
'shower': 6,  
'beforehand.': 1,

'cause': 16,  
'prob': 11,  
'u?\\\"': 1,  
'ham\"': 2,  
'anymore.': 4,  
'nothing': 15,  
'else': 13,  
'say.': 1,  
'Okay': 7,  
'name': 30,  
'price': 14,  
'long': 31,  
'legal!': 1,  
'Wen': 9,  
'them': 69,  
'up?': 9,  
'Y': 17,  
'ave': 4,  
'ams': 1,  
'xx': 7,  
'buy.': 3,  
'And': 98,  
'gone': 13,  
'4the': 1,  
'driving': 20,  
'test': 16,  
'yet.': 11,  
'wow.': 1,  
'You're': 9,  
'right!': 3,  
'mean': 18,  
'guess': 22,  
'gave': 10,  
'boston': 3,  
'men': 6,  
'changed': 5,  
'search': 14,  
'location': 2,  
'nyc,': 1,  
'something': 47,  
'changed.': 1,  
'Cuz': 2,  
'signin': 1,  
'page': 5,  
'says': 26,  
'boston.': 1,  
'Umma': 1,

```

'life': 35,
'vava': 3,
'umma': 5,
'lot': 26,
'dear': 36,
'wishes': 8,
'birthday.': 6,
'making': 21,
'birthday': 15,
'truly': 4,
'memorable.': 1,
'Aight,': 15,
'hit': 11,
'would': 66,
'ip': 1,
'address': 13,
'considering': 3,
'computer': 10,
"isn't": 9,
'minecraft': 1,
'server': 1,
'know!': 8,
'Grumpy': 1,
'old': 15,
...}

```

```

[53]: top_15_word = pd.Series(corpus.split()).value_counts().head(15)
      top_15_word

```

```

[53]: to      2134
      you     1622
      I       1466
      a       1327
      the     1197
      and      857
      in       798
      is       781
      i        742
      u        692
      for      644
      my       628
      of       591
      your     560
      me       557
      dtype: int64

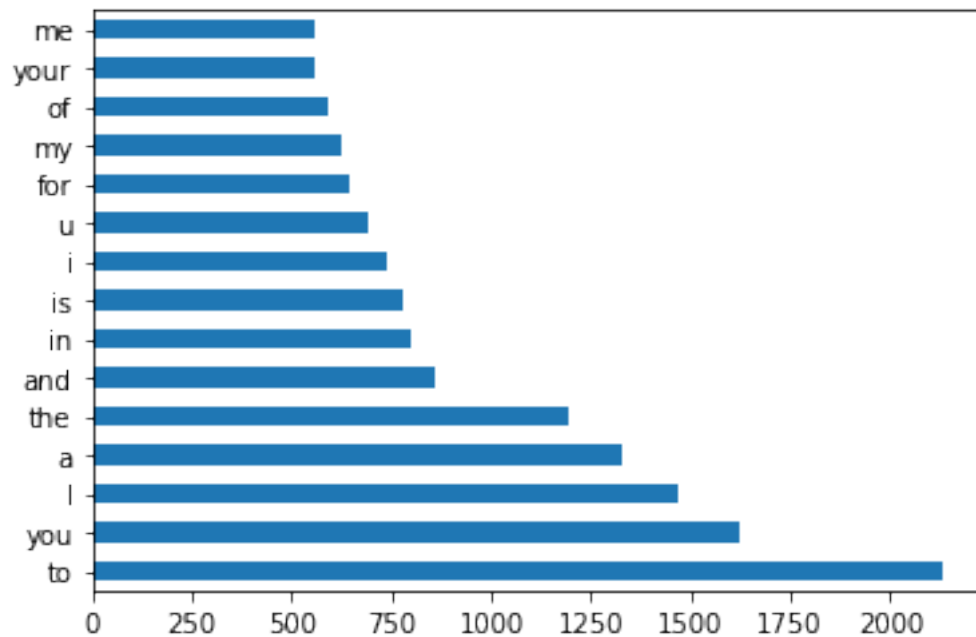
```

```

[54]: top_15_word.plot(kind = "barh")

```

```
[54]: <matplotlib.axes._subplots.AxesSubplot at 0x7f097457a110>
```



```
[55]: pd.Series(corpus.split()).value_counts()
```

```
[55]: to          2134
      you        1622
      I          1466
      a          1327
      the        1197
      ...
      page.       1
      DIDN'T      1
      rupaul       1
      TOOL!        1
      Rofl.        1
      Length: 15583, dtype: int64
```

```
[56]: pd.Series(corpus.split()).nunique()
```

```
[56]: 15583
```

```
[57]: X = dataset.v2
      Y = dataset.v1
      encoder = LabelEncoder()
      Y = encoder.fit_transform(Y)
      Y = Y.reshape(-1,1)
```



```
[58]: X_train,X_test,y_train,y_test = train_test_split(X,Y,test_size=0.2)
tokenizer = Tokenizer(num_words=2000,lower=True)
tokenizer.fit_on_texts(X_train)
sequences = tokenizer.texts_to_sequences(X_train)
X_train=sequence.pad_sequences(sequences,maxlen=200)
```

## 2 Create Model

```
[59]: model = Sequential()
```

## 3 Add Layers(LSTM,Dense,Hidden-Layers,Output)

```
[60]: model.add(Embedding(2000,50,input_length=200))
model.add(LSTM(64))
model.add(Dense(256,activation="relu"))
model.add(Dropout(0.5))
model.add(Dense(1,activation="sigmoid"))
```

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

Model: "sequential"

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 200, 50)	100000
lstm (LSTM)	(None, 64)	29440
dense (Dense)	(None, 256)	16640
dropout (Dropout)	(None, 256)	0
dense_1 (Dense)	(None, 1)	257

```
=====  
Total params: 146,337  
Trainable params: 146,337  
Non-trainable params: 0  
-----
```

## 4 Compile the Model

```
[62]: model.  
      ↪ compile(loss="binary_crossentropy", optimizer=RMSprop(), metrics=['accuracy'])
```

## 5 Fit the Model

```
[63]: model.fit(X_train, y_train, batch_size=128, epochs=20, validation_split=0.2)
```

```
Epoch 1/20  
28/28 [=====] - 8s 36ms/step - loss: 0.3125 - accuracy:  
0.8898 - val_loss: 0.1360 - val_accuracy: 0.9742  
Epoch 2/20  
28/28 [=====] - 0s 16ms/step - loss: 0.0829 - accuracy:  
0.9804 - val_loss: 0.0509 - val_accuracy: 0.9854  
Epoch 3/20  
28/28 [=====] - 0s 17ms/step - loss: 0.0353 - accuracy:  
0.9910 - val_loss: 0.0546 - val_accuracy: 0.9854  
Epoch 4/20  
28/28 [=====] - 0s 16ms/step - loss: 0.0253 - accuracy:  
0.9927 - val_loss: 0.0396 - val_accuracy: 0.9888  
Epoch 5/20  
28/28 [=====] - 0s 16ms/step - loss: 0.0130 - accuracy:  
0.9961 - val_loss: 0.0759 - val_accuracy: 0.9865  
Epoch 6/20  
28/28 [=====] - 0s 17ms/step - loss: 0.0119 - accuracy:  
0.9964 - val_loss: 0.0518 - val_accuracy: 0.9910  
Epoch 7/20  
28/28 [=====] - 1s 35ms/step - loss: 0.0088 - accuracy:  
0.9975 - val_loss: 0.0567 - val_accuracy: 0.9888  
Epoch 8/20  
28/28 [=====] - 0s 17ms/step - loss: 0.0039 - accuracy:  
0.9989 - val_loss: 0.0798 - val_accuracy: 0.9798  
Epoch 9/20  
28/28 [=====] - 0s 16ms/step - loss: 0.0032 - accuracy:  
0.9994 - val_loss: 0.0858 - val_accuracy: 0.9910  
Epoch 10/20  
28/28 [=====] - 0s 17ms/step - loss: 0.0034 - accuracy:  
0.9989 - val_loss: 0.0749 - val_accuracy: 0.9910  
Epoch 11/20  
28/28 [=====] - 0s 17ms/step - loss: 0.0024 - accuracy:  
0.9994 - val_loss: 0.0848 - val_accuracy: 0.9910  
Epoch 12/20  
28/28 [=====] - 0s 16ms/step - loss: 0.0013 - accuracy:  
0.9997 - val_loss: 0.0917 - val_accuracy: 0.9888
```

```

Epoch 13/20
28/28 [=====] - 0s 18ms/step - loss: 7.5109e-04 -
accuracy: 0.9997 - val_loss: 0.1021 - val_accuracy: 0.9910
Epoch 14/20
28/28 [=====] - 0s 17ms/step - loss: 4.7842e-04 -
accuracy: 1.0000 - val_loss: 0.1063 - val_accuracy: 0.9865
Epoch 15/20
28/28 [=====] - 0s 17ms/step - loss: 9.3035e-04 -
accuracy: 0.9997 - val_loss: 0.1122 - val_accuracy: 0.9910
Epoch 16/20
28/28 [=====] - 0s 17ms/step - loss: 0.0011 - accuracy:
0.9997 - val_loss: 0.1114 - val_accuracy: 0.9865
Epoch 17/20
28/28 [=====] - 0s 17ms/step - loss: 1.2910e-04 -
accuracy: 1.0000 - val_loss: 0.1164 - val_accuracy: 0.9899
Epoch 18/20
28/28 [=====] - 0s 17ms/step - loss: 4.6087e-05 -
accuracy: 1.0000 - val_loss: 0.1303 - val_accuracy: 0.9899
Epoch 19/20
28/28 [=====] - 0s 17ms/step - loss: 1.1798e-05 -
accuracy: 1.0000 - val_loss: 0.1497 - val_accuracy: 0.9899
Epoch 20/20
28/28 [=====] - 0s 17ms/step - loss: 3.5899e-06 -
accuracy: 1.0000 - val_loss: 0.1707 - val_accuracy: 0.9899

```

[63]: <keras.callbacks.History at 0x7f0974786510>

## 6 Save the Model

```
[64]: model.save("spam_Model.h5")
```

## 7 Test the Model

```
[65]: max_len = 150
test_sequences = tokenizer.texts_to_sequences(X_test)
test_sequences_matrix = keras.utils.pad_sequences(test_sequences,maxlen=max_len)
acc = model.evaluate(test_sequences_matrix,y_test)
```

```

35/35 [=====] - 1s 8ms/step - loss: 0.2493 - accuracy:
0.9794

```

```
[66]: def predict(message):
        t = tokenizer.texts_to_sequences(message)
        t = sequence.pad_sequences(t, maxlen=200)
```

```
predicts = model.predict(t)
if predicts > 0.5:
    print("Spam")
else:
    print("Not Spam")
```

### 7.0.1 Test 1 : Not Spam Message

```
[70]: predict(["Today Class for you!!"])
```

```
1/1 [=====] - 0s 21ms/step
Not Spam
```

### 7.0.2 Test 2 : Spam Message

```
[69]: predict(["You have offer for buying new mobile"])
```

```
1/1 [=====] - 0s 21ms/step
Spam
```