Importing the libraries

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from keras.models import Model,Sequential
from keras.layers import LSTM, Activation, Dense, Dropout, Input,
Embedding
from keras.optimizers import Adam
from keras.preprocessing.text import Tokenizer
from keras.preprocessing import sequence
from keras.utils import pad_sequences
from keras.utils import to_categorical
from keras.callbacks import EarlyStopping
%matplotlib inline

from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

Reading the dataset

```python
df = pd.read_csv('/content/drive/MyDrive/spam.csv',encoding='latin-1')
df.head()
```

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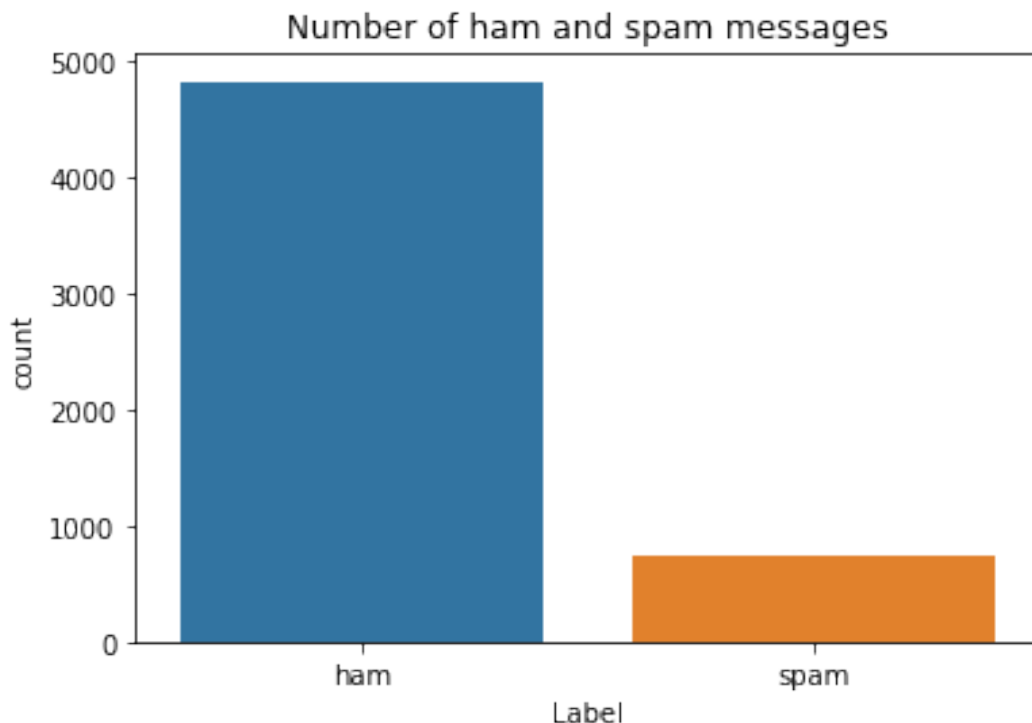
```
df.drop(['Unnamed: 2', 'Unnamed: 3', 'Unnamed:
4'],axis=1,inplace=True)
df.info()
```

```
<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
```

```
sns.countplot(df.v1)
plt.xlabel('Label')
plt.title('Number of ham and spam messages')
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43:
FutureWarning: Pass the following variable as a keyword arg: x. From
version 0.12, the only valid positional argument will be `data`, and
passing other arguments without an explicit keyword will result in an
error or misinterpretation.
  FutureWarning
```

```
Text(0.5, 1.0, 'Number of ham and spam messages')
```

```python
X = df.v2
Y = df.v1
le = LabelEncoder()
Y = le.fit_transform(Y)
Y = Y.reshape(-1,1)

X_train,X_test,Y_train,Y_test = train_test_split(X,Y,test_size=0.15)
```

Process the data

```python
max_words = 1000
max_len = 150
tok = Tokenizer(num_words=max_words)
tok.fit_on_texts(X_train)
sequences = tok.texts_to_sequences(X_train)
sequences_matrix = pad_sequences(sequences,maxlen=max_len)
sequences_matrix
```

```
array([[  0,   0,   0, ...,  15,   2,  27],
       [  0,   0,   0, ..., 260,  10,   2],
       [  0,   0,   0, ...,  13, 222,  23],
       ...,
       [  0,   0,   0, ..., 259,   9,  11],
       [  0,   0,   0, ...,  16,  10, 353],
       [  0,   0,   0, ..., 901, 200, 196]], dtype=int32)
```

```python
tok.index_word
```

```
{1: 'i',
 2: 'to',
 3: 'you',
 4: 'a',
 5: 'the',
 6: 'u',
 7: 'and',
 8: 'in',
 9: 'is',
 10: 'me',
 11: 'my',
 12: 'for',
 13: 'your',
 14: 'it',
 15: 'of',
 16: 'call',
 17: 'have',
 18: 'on',
 19: 'that',
 20: 'now',
 21: '2',
 22: 'are',
 23: 'so',
```

```
24: 'not',
25: 'or',
26: 'but',
27: 'do',
28: 'can',
29: 'ur',
30: 'at',
31: 'get',
32: "i'm",
33: 'if',
34: 'will',
35: 'be',
36: 'just',
37: 'with',
38: 'no',
39: 'we',
40: 'this',
41: '4',
42: 'up',
43: 'gt',
44: 'lt',
45: 'ok',
46: 'when',
47: 'from',
48: 'free',
49: 'how',
50: 'all',
51: 'go',
52: 'know',
53: 'out',
54: 'what',
55: 'like',
56: 'then',
57: 'got',
58: 'come',
59: 'good',
60: 'was',
61: 'am',
62: 'only',
63: 'love',
64: 'its',
65: 'day',
66: 'time',
67: 'send',
68: 'text',
69: 'he',
70: 'there',
71: 'want',
72: 'as',
73: 'txt',
```

```
74:  'by',
75:  'going',
76:  'one',
77:  'home',
78:  'about',
79:  'sorry',
80:  "i'll",
81:  'lor',
82:  'need',
83:  'r',
84:  'still',
85:  'da',
86:  'dont',
87:  'back',
88:  'stop',
89:  'n',
90:  'today',
91:  'see',
92:  'mobile',
93:  'our',
94:  'reply',
95:  "don't",
96:  'take',
97:  'tell',
98:  'pls',
99:  'k',
100: 'please',
101: 'she',
102: 'later',
103: 'hi',
104: 'think',
105: 'any',
106: 'been',
107: '1',
108: 'ì',
109: 'dear',
110: 'they',
111: 'her',
112: 'much',
113: 'phone',
114: 'new',
115: 'week',
116: 'hope',
117: 'did',
118: 'who',
119: 'here',
120: 'night',
121: 'some',
122: 'msg',
123: 'has',
```

```
124: 'claim',
125: 'an',
126: 'well',
127: 'great',
128: 'yes',
129: 'too',
130: 'd',
131: 'where',
132: 'had',
133: 'make',
134: 'more',
135: 'e',
136: 'oh',
137: 'hey',
138: "it's",
139: 'message',
140: 'him',
141: 'way',
142: 'number',
143: 'c',
144: 'tomorrow',
145: 'happy',
146: 'say',
147: 'wat',
148: 'work',
149: 'give',
150: 'right',
151: 'should',
152: 'amp',
153: 'www',
154: 'doing',
155: 'prize',
156: 'already',
157: 'after',
158: 'ask',
159: 'cash',
160: 'them',
161: 'said',
162: 'im',
163: '3',
164: 'why',
165: 'meet',
166: 'miss',
167: 'yeah',
168: 'last',
169: 'really',
170: 'very',
171: 't',
172: 'life',
173: 'would',
```

```
174: 'morning',
175: 'babe',
176: 'find',
177: 'b',
178: 'cos',
179: 'every',
180: 'thanks',
181: 'also',
182: '150p',
183: 'urgent',
184: 'contact',
185: 'pick',
186: 'nokia',
187: 'anything',
188: 'win',
189: 'sent',
190: 'sure',
191: 'let',
192: 'won',
193: 'buy',
194: 'care',
195: 'lol',
196: 'min',
197: "i've",
198: 'keep',
199: 'com',
200: '5',
201: 'cant',
202: 'wait',
203: 'over',
204: 'uk',
205: 'off',
206: 'something',
207: 'before',
208: 'around',
209: 'even',
210: '50',
211: 'many',
212: 'tone',
213: 'first',
214: 'thing',
215: 'could',
216: 'next',
217: 'which',
218: 'feel',
219: 'nice',
220: 'again',
221: 'leave',
222: 'always',
223: "can't",
```

```
224: 'were',
225: 'late',
226: 'sleep',
227: 'gonna',
228: 'place',
229: 'chat',
230: 'his',
231: 'soon',
232: 'us',
233: 'per',
234: "that's",
235: 'co',
236: 'money',
237: 'customer',
238: 'service',
239: 'sms',
240: 'x',
241: 'down',
242: 'friends',
243: 'ya',
244: 'dun',
245: '16',
246: 'wan',
247: 'things',
248: 'v',
249: 'someone',
250: 'help',
251: 'hello',
252: 's',
253: 'waiting',
254: 'coming',
255: 'went',
256: 'ìï',
257: 'tonight',
258: 'getting',
259: 'wish',
260: 'told',
261: 'fine',
262: 'same',
263: 'name',
264: 'special',
265: 'thk',
266: 'best',
267: 'guaranteed',
268: 'friend',
269: 'year',
270: 'heart',
271: 'gud',
272: 'holiday',
273: 'done',
```

```
274: '18',
275: 'mins',
276: 'stuff',
277: 'may',
278: 'cs',
279: 'thought',
280: 'class',
281: 'other',
282: 'yet',
283: "you're",
284: '7',
285: 'draw',
286: 'thats',
287: 'talk',
288: 'try',
289: 'people',
290: 'having',
291: 'line',
292: 'haha',
293: '6',
294: 'man',
295: 'never',
296: 'person',
297: 'yup',
298: 'lot',
299: 'better',
300: 'lunch',
301: 'long',
302: 'meeting',
303: 'being',
304: 'cool',
305: 'trying',
306: 'few',
307: 'y',
308: '1st',
309: 'days',
310: 'use',
311: 'mind',
312: 'live',
313: 'latest',
314: 'camera',
315: 'job',
316: 'ill',
317: 'wk',
318: 'dat',
319: 'check',
320: 'god',
321: 'big',
322: 'car',
323: 'ready',
```

```
324: 'guys',
325: "didn't",
326: 'because',
327: 'chance',
328: 'real',
329: 'awarded',
330: 'shit',
331: 'enjoy',
332: 'shows',
333: 'liao',
334: 'girl',
335: 'finish',
336: 'bit',
337: 'receive',
338: 'å£1',
339: 'wanna',
340: 'smile',
341: 'lar',
342: 'end',
343: 'guess',
344: 'start',
345: 'half',
346: 'house',
347: 'world',
348: '150ppm',
349: 'word',
350: 'sweet',
351: 'dinner',
352: 'landline',
353: 'sir',
354: 'account',
355: 'play',
356: 'box',
357: 'problem',
358: 'watch',
359: 'another',
360: 'into',
361: 'offer',
362: 'luv',
363: 'birthday',
364: 'quite',
365: '10',
366: 'dis',
367: 'cost',
368: 'might',
369: 'video',
370: 'eat',
371: 'than',
372: 'xxx',
373: 'apply',
```

```
374: 'ever',
375: 'remember',
376: 'rate',
377: 'sat',
378: '9',
379: 'tv',
380: 'left',
381: 'thanx',
382: 'early',
383: 'plan',
384: 'code',
385: 'princess',
386: 'does',
387: 'kiss',
388: 'jus',
389: 'month',
390: 'po',
391: 'ringtone',
392: 'part',
393: 'forgot',
394: 'll',
395: 'bt',
396: 'looking',
397: 'yo',
398: 'orange',
399: 'didnt',
400: 'speak',
401: 'baby',
402: 'bed',
403: 'actually',
404: 'nothing',
405: 'probably',
406: 'once',
407: 'two',
408: 'leh',
409: 'dunno',
410: 'ah',
411: 'hear',
412: 'fun',
413: 'room',
414: 'pay',
415: "how's",
416: 'maybe',
417: 'look',
418: 'aight',
419: 'update',
420: 'enough',
421: 'face',
422: 'å£1000',
423: 'award',
```

```
424: 'minutes',
425: 'texts',
426: 'came',
427: 'town',
428: 'network',
429: 'reach',
430: 'weekend',
431: 'collection',
432: '8',
433: 'wont',
434: 'watching',
435: 'pa',
436: 'true',
437: 'g',
438: 'called',
439: 'thank',
440: 'hour',
441: 'boy',
442: 'put',
443: 'since',
444: 'while',
445: '000',
446: 'everything',
447: '2nd',
448: 'bad',
449: 'between',
450: 'fuck',
451: 'm',
452: "he's",
453: 'school',
454: 'sexy',
455: 'join',
456: 'wen',
457: 'says',
458: 'though',
459: 'asked',
460: 'selected',
461: 'those',
462: 'bring',
463: 'evening',
464: 'times',
465: 'price',
466: 'until',
467: 'valid',
468: 'details',
469: 'tmr',
470: 'most',
471: 'entry',
472: 'dad',
473: 'shopping',
```

```
474: 'missing',
475: 'shall',
476: 'pain',
477: 'den',
478: 'goes',
479: 'hav',
480: 'away',
481: 'plz',
482: 'xmas',
483: 'easy',
484: 'anyway',
485: 'working',
486: 'office',
487: 'wif',
488: 'wake',
489: 'made',
490: 'afternoon',
491: 'able',
492: 'mail',
493: 'little',
494: 'de',
495: 'top',
496: 'yourself',
497: 'plus',
498: 'else',
499: 'music',
500: 'okay',
501: 'wife',
502: 'gift',
503: 'saw',
504: 'mob',
505: 'attempt',
506: 'driving',
507: 'double',
508: 'yours',
509: "we're",
510: 'till',
511: 'bored',
512: "there's",
513: 'id',
514: 'show',
515: 'must',
516: 'å£100',
517: 'nite',
518: 'test',
519: 'collect',
520: 'bus',
521: 're',
522: 'order',
523: 'sae',
```

```
524: 'wants',
525: 'means',
526: 'ard',
527: 'run',
528: 'coz',
529: 'messages',
530: "won't",
531: 'worry',
532: 'important',
533: 'abt',
534: 'colour',
535: 'calls',
536: 'tones',
537: '10p',
538: 'wot',
539: 'wanted',
540: 'weekly',
541: 'making',
542: 'todays',
543: 'alright',
544: 'guy',
545: 'yesterday',
546: 'family',
547: 'hair',
548: 'decimal',
549: 'juz',
550: 'sch',
551: 'await',
552: 'dude',
553: 'without',
554: 'wid',
555: 'beautiful',
556: 'national',
557: 'full',
558: 'saying',
559: 'hours',
560: 'til',
561: '8007',
562: 'hurt',
563: 'date',
564: 'tot',
565: 'havent',
566: 'words',
567: 'pm',
568: 'book',
569: 'net',
570: 'congrats',
571: 'drive',
572: 'lei',
573: 'calling',
```

```
574: 'happen',
575: 'feeling',
576: 'comes',
577: 'old',
578: 'missed',
579: 'email',
580: 'pub',
581: 'auction',
582: 'points',
583: 'ring',
584: 'row',
585: 'noe',
586: 'friendship',
587: 'sleeping',
588: 'aft',
589: '500',
590: 'question',
591: 'address',
592: 'pic',
593: 'final',
594: 'charge',
595: 'years',
596: 'tried',
597: 'mean',
598: 'movie',
599: 'stay',
600: 'land',
601: 'busy',
602: 'brother',
603: 'games',
604: 'food',
605: 'tomo',
606: 'makes',
607: 'online',
608: 'second',
609: 'everyone',
610: 'both',
611: 'answer',
612: 'å£2000',
613: "haven't",
614: 'story',
615: 'voucher',
616: 'either',
617: 'walk',
618: 'dreams',
619: 'trip',
620: '86688',
621: 'post',
622: 'news',
623: 'wil',
```

```
624: 'leaving',
625: 'vouchers',
626: 'private',
627: 'dating',
628: 'together',
629: 'å£5000',
630: 'these',
631: 'started',
632: '12hrs',
633: 'haf',
634: 'change',
635: 'pics',
636: 'eve',
637: 'shop',
638: 'card',
639: 'believe',
640: 'hot',
641: 'oso',
642: 'goin',
643: 'rite',
644: 'chikku',
645: 'å£500',
646: 'sad',
647: 'choose',
648: 'tho',
649: 'break',
650: 'don',
651: 'alone',
652: 'bonus',
653: 'search',
654: 'club',
655: "c's",
656: 'prob',
657: 'okie',
658: 'lose',
659: 'statement',
660: 'close',
661: 'game',
662: 'available',
663: 'anyone',
664: 'expires',
665: 'pobox',
666: 'finished',
667: 'angry',
668: 'ha',
669: 'kind',
670: 'frnd',
671: '11',
672: 'cause',
673: 'needs',
```

```
674: 'worth',
675: '100',
676: '08000930705',
677: 'used',
678: 'goodmorning',
679: 'set',
680: 'xx',
681: 'mine',
682: 'sounds',
683: 'unsubscribe',
684: 'minute',
685: 'forget',
686: '750',
687: 'anytime',
688: 'delivery',
689: 'suite342',
690: '2lands',
691: 'sis',
692: 'huh',
693: "she's",
694: 'pounds',
695: 'lets',
696: 'happened',
697: 'drink',
698: "what's",
699: 'touch',
700: 'hand',
701: 'services',
702: '30',
703: 'identifier',
704: 'knw',
705: 'saturday',
706: 'sister',
707: 'gd',
708: "'''",
709: 'mates',
710: 'hard',
711: 'neva',
712: 'wonderful',
713: 'lots',
714: 'http',
715: 'chennai',
716: 'wit',
717: 'company',
718: 'sun',
719: 'listen',
720: 'å£3',
721: 'lovely',
722: 'far',
723: 'took',
```

```
724: 'taking',
725: 'light',
726: 'meant',
727: 'camcorder',
728: 'gone',
729: 'parents',
730: 'pretty',
731: 'gr8',
732: "you've",
733: 'sunday',
734: 'content',
735: 'head',
736: 'thinks',
737: 'savamob',
738: 'smth',
739: 'whatever',
740: 'o',
741: 'sea',
742: 'boytoy',
743: 'treat',
744: 'friday',
745: 'opt',
746: 'whats',
747: 'oredi',
748: 'finally',
749: 'visit',
750: 'takes',
751: 'gas',
752: 'wrong',
753: 'crazy',
754: 'side',
755: 'loving',
756: "'",
757: 'drop',
758: 'girls',
759: 'quiz',
760: 'congratulations',
761: 'fr',
762: 'earlier',
763: 'fancy',
764: 'each',
765: 'hold',
766: 'phones',
767: 'tel',
768: 'tired',
769: '08000839402',
770: 'nope',
771: 'offers',
772: 'darlin',
773: 'type',
```

```
774: "we'll",
775: 'etc',
776: '2003',
777: 'secret',
778: 'wkly',
779: 'fri',
780: 'hows',
781: 'bslvyl',
782: 'hmm',
783: 'w',
784: 'mobileupd8',
785: '800',
786: 'un',
787: 'thinking',
788: 'smiling',
789: 'their',
790: 'ten',
791: 'rs',
792: 'outside',
793: 'f',
794: 'weeks',
795: 'smoke',
796: 'ltd',
797: 'unlimited',
798: 'bout',
799: 'hmmm',
800: 'winner',
801: 'frnds',
802: 'press',
803: 'through',
804: 'talking',
805: 'charged',
806: 'blue',
807: 'rental',
808: 'cut',
809: 'download',
810: 'nt',
811: 'msgs',
812: 'balance',
813: 'currently',
814: 'reading',
815: 'uncle',
816: 'college',
817: 'india',
818: "you'll",
819: 'fone',
820: 'fucking',
821: 'mate',
822: 'valued',
823: 'complimentary',
```

```
824: 'hee',
825: 'march',
826: 'ends',
827: 'reveal',
828: 'stupid',
829: 'å£250',
830: 'redeemed',
831: '04',
832: '20',
833: 'lesson',
834: 'tc',
835: 'dnt',
836: 'fast',
837: 'mayb',
838: 'awesome',
839: 'credit',
840: 'support',
841: 'happiness',
842: 'mobiles',
843: 'bath',
844: '\x89û',
845: 'gets',
846: 'almost',
847: 'whole',
848: 'couple',
849: 'christmas',
850: 'enter',
851: '87066',
852: 'park',
853: 'laptop',
854: 'yar',
855: 'least',
856: 'luck',
857: 'via',
858: 'cum',
859: 'numbers',
860: 'å£350',
861: 'lost',
862: 'gn',
863: 'hg',
864: 'hit',
865: 'poly',
866: 'mum',
867: 'loads',
868: 'info',
869: 'player',
870: 'caller',
871: '03',
872: 'call2optout',
873: 'read',
```

874: 'ago',
875: 'case',
876: 'met',
877: '12',
878: 'freemsg',
879: 'lovable',
880: 'w1j6hl',
881: 'comp',
882: 'bcoz',
883: 'die',
884: 'song',
885: 'open',
886: 'remove',
887: 'mom',
888: 'age',
889: 'darren',
890: 'st',
891: 'carlos',
892: 'somebody',
893: 'project',
894: 'sending',
895: 'frm',
896: 'slowly',
897: '0800',
898: 'direct',
899: 'found',
900: 'welcome',
901: 'within',
902: 'mu',
903: 'txting',
904: 'txts',
905: 'simple',
906: 'pass',
907: 'lucky',
908: 'months',
909: 'checking',
910: 'seen',
911: 'ni8',
912: 'operator',
913: 'nyt',
914: "doesn't",
915: 'callertune',
916: 'booked',
917: 'pray',
918: 'seeing',
919: 'å£10',
920: 'picking',
921: 'yep',
922: 'å£2',
923: 'gave',

```
924: 'correct',
925: 'ans',
926: 'hoping',
927: 'b4',
928: 'telling',
929: 'comin',
930: 'area',
931: 'sony',
932: 'bank',
933: 'pete',
934: "wasn't",
935: '00',
936: 'computer',
937: 'discount',
938: 'std',
939: 'ldn',
940: 'normal',
941: 'gym',
942: 'somewhere',
943: 'john',
944: 'understand',
945: 'kate',
946: 'admirer',
947: 'mrng',
948: 'red',
949: 'write',
950: 'semester',
951: 'paper',
952: 'jay',
953: 'felt',
954: 'sex',
955: 'sound',
956: 'reward',
957: 'point',
958: 'cheap',
959: 'wana',
960: 'del',
961: 'blood',
962: 'wap',
963: 'motorola',
964: 'shower',
965: 'na',
966: 'monday',
967: 'reason',
968: 'save',
969: 'course',
970: 'confirm',
971: 'across',
972: 'wine',
973: 'lazy',
```

```
 974: 'mode',
 975: 'less',
 976: 'silent',
 977: 'truth',
 978: 'safe',
 979: 'own',
 980: 'swing',
 981: 'heard',
 982: "isn't",
 983: 'dead',
 984: 'copy',
 985: 'store',
 986: 'ts',
 987: 'ldew',
 988: 'accept',
 989: 'knew',
 990: 'water',
 991: 'entered',
 992: 'usual',
 993: 'supposed',
 994: 'doesnt',
 995: 'abiola',
 996: 'ipod',
 997: 'asap',
 998: 'sofa',
 999: 'short',
 1000: 'poor',
 ...}
```

Creating model and Adding the layers

```python
TOT_SIZE = len(tok.word_index)+1
lstm_model = Sequential()
lstm_model.add(Embedding(TOT_SIZE, 32, input_length=max_len))
lstm_model.add(LSTM(100))
lstm_model.add(Dropout(0.4))
lstm_model.add(Dense(20, activation="relu"))
lstm_model.add(Dropout(0.3))
lstm_model.add(Dense(1, activation = "sigmoid"))
```

Compile the model

```python
lstm_model.compile(loss = "binary_crossentropy", optimizer = "adam",
metrics = ["accuracy"])
lstm_model.summary()
```

```
Model: "sequential"
```

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 150, 32)	262304

```
lstm (LSTM)                  (None, 100)              53200

dropout (Dropout)            (None, 100)              0

dense (Dense)                (None, 20)               2020

dropout_1 (Dropout)          (None, 20)               0

dense_1 (Dense)              (None, 1)                21

=================================================================
Total params: 317,545
Trainable params: 317,545
Non-trainable params: 0
_____
```

Fit the model

```
lstm_model.fit(sequences_matrix,Y_train,batch_size=128,epochs=10,
          validation_split=0.2,
          workers=10,

callbacks=[EarlyStopping(monitor='val_loss',min_delta=0.0001)])

Epoch 1/10
30/30 [==============================] - 14s 336ms/step - loss: 0.4534
- accuracy: 0.8508 - val_loss: 0.2814 - val_accuracy: 0.8787
Epoch 2/10
30/30 [==============================] - 6s 212ms/step - loss: 0.1870
- accuracy: 0.9388 - val_loss: 0.1044 - val_accuracy: 0.9757

<keras.callbacks.History at 0x7fd849a023d0>
```

Save the model

```
lstm_model.save('sms.h5')
```

Test the model

```
test_sequences = tok.texts_to_sequences(X_test)
test_sequences_matrix = pad_sequences(test_sequences,maxlen=max_len)

acc = lstm_model.evaluate(test_sequences_matrix,Y_test)

27/27 [==============================] - 1s 20ms/step - loss: 0.0805 -
accuracy: 0.9880

print('Test set\n  Loss: {:0.3f}\n  Accuracy:
{:0.3f}'.format(acc[0],acc[1]))
```

```
Test set
  Loss: 0.081
  Accuracy: 0.988
```