

Importing the libraries

```
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
```

Reading the dataset

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

```
   v1  ... Unnamed: 4
0  ham  ...        NaN
1  ham  ...        NaN
2 spam  ...        NaN
3  ham  ...        NaN
4  ham  ...        NaN
```

[5 rows x 5 columns]

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

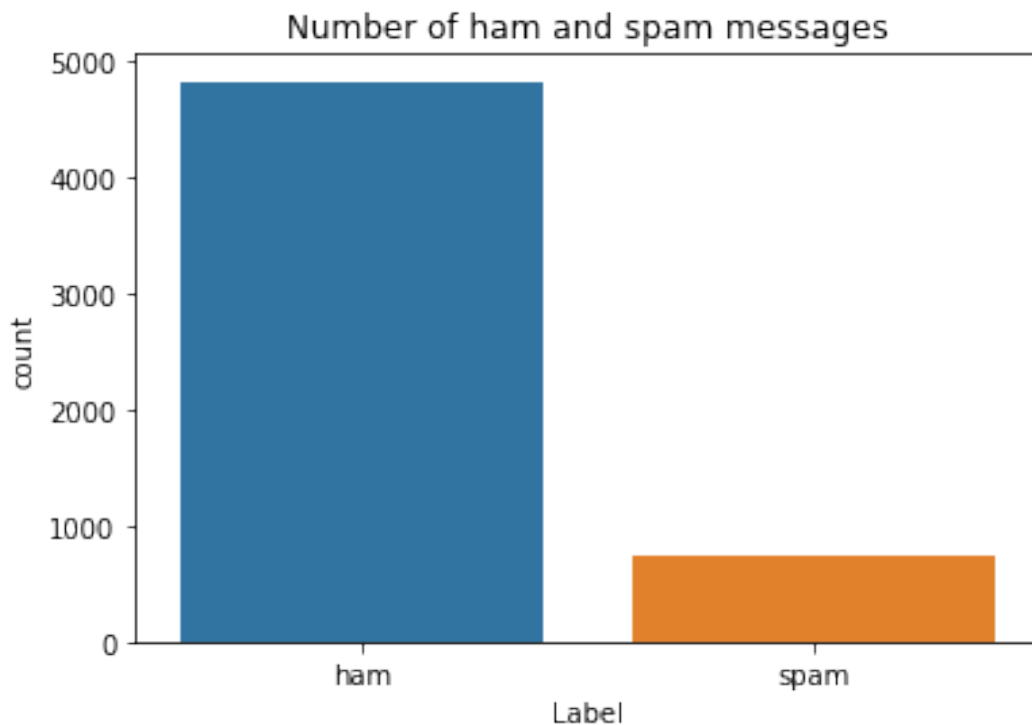
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5572 entries, 0 to 5571
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype
---  -
 0    v1      5572 non-null    object
 1    v2      5572 non-null    object
dtypes: object(2)
memory usage: 87.2+ KB
```

```
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')
```



```
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

```
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, ..., 135, 453, 391],
       [ 0,  0,  0, ..., 65, 782, 842],
```

```

[ 0, 0, 0, ..., 82, 2, 176],
...,
[ 0, 0, 0, ..., 51, 960, 256],
[ 0, 0, 0, ..., 49, 12, 298],
[ 0, 0, 0, ..., 753, 853, 59]], dtype=int32)

```

```
tok.index_word
```

```

{1: 'i',
 2: 'to',
 3: 'you',
 4: 'a',
 5: 'the',
 6: 'u',
 7: 'and',
 8: 'is',
 9: 'in',
10: 'me',
11: 'my',
12: 'for',
13: 'your',
14: 'it',
15: 'of',
16: 'call',
17: 'have',
18: 'on',
19: '2',
20: 'that',
21: 'are',
22: 'now',
23: 'so',
24: 'not',
25: 'or',
26: 'but',
27: 'do',
28: 'at',
29: 'get',
30: 'can',
31: 'ur',
32: 'if',
33: 'will',
34: 'be',
35: "i'm",
36: 'no',
37: 'just',
38: 'with',
39: 'we',
40: 'this',
41: '4',
42: 'gt',
43: 'up',

```

44: 'lt',
45: 'how',
46: 'when',
47: 'free',
48: 'ok',
49: 'go',
50: 'all',
51: 'from',
52: 'out',
53: 'good',
54: 'got',
55: 'know',
56: 'was',
57: 'what',
58: 'then',
59: 'like',
60: 'its',
61: 'day',
62: 'am',
63: 'time',
64: 'come',
65: 'only',
66: 'love',
67: 'there',
68: 'text',
69: 'he',
70: 'send',
71: 'want',
72: 'txt',
73: 'going',
74: 'one',
75: 'need',
76: 'as',
77: 'home',
78: 'by',
79: "i'll",
80: 'about',
81: 'sorry',
82: 'stop',
83: 'see',
84: 'r',
85: 'still',
86: 'lor',
87: 'n',
88: 'back',
89: 'today',
90: 'reply',
91: 'tell',
92: 'our',
93: 'da',

94: 'she',
95: 'dont',
96: 'hi',
97: 'new',
98: 'take',
99: "don't",
100: 'mobile',
101: 'dear',
102: 'pls',
103: 'please',
104: 'k',
105: 'any',
106: 'later',
107: 'think',
108: 'her',
109: 'did',
110: 'some',
111: 'an',
112: 'been',
113: 'much',
114: 'week',
115: 'great',
116: 'they',
117: 'phone',
118: 'i',
119: 'oh',
120: 'claim',
121: 'well',
122: 'here',
123: 'l',
124: 'who',
125: 'him',
126: 'has',
127: 'd',
128: 'more',
129: 'hey',
130: 'had',
131: 'hope',
132: 'where',
133: 'too',
134: 'night',
135: 'msg',
136: 'happy',
137: 'www',
138: 'number',
139: 'c',
140: 'should',
141: 'wat',
142: 'make',
143: 'give',

144: 'prize',
145: 'yes',
146: 'message',
147: 'work',
148: 'after',
149: 'really',
150: 'tomorrow',
151: "it's",
152: '3',
153: 'e',
154: 'cash',
155: 'said',
156: 'amp',
157: 'way',
158: 'yeah',
159: 'say',
160: 'find',
161: 'already',
162: 'doing',
163: 'ask',
164: 'why',
165: 'im',
166: 'right',
167: 'babe',
168: 'life',
169: 'them',
170: 'cos',
171: 'last',
172: 'very',
173: 'would',
174: 'win',
175: 'lol',
176: '150p',
177: 'miss',
178: 'urgent',
179: 't',
180: 'pick',
181: 'morning',
182: 'won',
183: 'every',
184: 'thanks',
185: 'sent',
186: 'meet',
187: 'something',
188: 'care',
189: 'nokia',
190: 'over',
191: 'us',
192: 'b',
193: 'let',

194: 'again',
195: 'also',
196: 'com',
197: 'anything',
198: 'buy',
199: 'contact',
200: 'first',
201: 'sure',
202: 'someone',
203: 'even',
204: 'uk',
205: 'thing',
206: 'soon',
207: 'tone',
208: 'which',
209: 'before',
210: 'keep',
211: "can't",
212: 'service',
213: 'wait',
214: 'cant',
215: 'his',
216: 's',
217: 'sleep',
218: 'around',
219: 'min',
220: 'went',
221: 'place',
222: 'next',
223: '5',
224: "i've",
225: 'money',
226: 'were',
227: 'leave',
228: 'feel',
229: 'could',
230: 'always',
231: 'co',
232: 'told',
233: 'nice',
234: 'many',
235: 'late',
236: 'gonna',
237: 'tonight',
238: 'wish',
239: '50',
240: 'year',
241: 'other',
242: 'friend',
243: 'x',

244: '16',
245: 'waiting',
246: 'per',
247: "that's",
248: 'down',
249: 'off',
250: 'customer',
251: 'chat',
252: 'ii',
253: 'hello',
254: 'ya',
255: 'getting',
256: 'help',
257: 'friends',
258: 'gud',
259: 'wan',
260: "you're",
261: 'dun',
262: 'fine',
263: 'sms',
264: 'name',
265: 'done',
266: '6',
267: 'guaranteed',
268: 'try',
269: 'yet',
270: '18',
271: 'same',
272: 'class',
273: 'haha',
274: 'special',
275: 'holiday',
276: 'mins',
277: 'smile',
278: 'people',
279: 'heart',
280: 'things',
281: '7',
282: 'cs',
283: 'best',
284: 'thk',
285: 'coming',
286: 'use',
287: 'stuff',
288: 'ill',
289: 'may',
290: 'cool',
291: 'thought',
292: 'finish',
293: "didn't",

294: 'bit',
295: 'y',
296: 'chance',
297: 'v',
298: 'lunch',
299: 'latest',
300: 'trying',
301: 'better',
302: 'man',
303: 'having',
304: 'days',
305: 'person',
306: 'ready',
307: 'because',
308: 'being',
309: 'thats',
310: 'meeting',
311: 'long',
312: 'eat',
313: 'yo',
314: 'problem',
315: 'talk',
316: '1st',
317: 'end',
318: 'guess',
319: 'house',
320: 'receive',
321: 'than',
322: 'live',
323: 'wk',
324: 'car',
325: 'lot',
326: 'into',
327: 'big',
328: 'line',
329: 'draw',
330: 'mind',
331: 'word',
332: 'never',
333: 'sir',
334: 'god',
335: 'half',
336: 'ever',
337: 'real',
338: 'yup',
339: 'few',
340: 'shit',
341: 'quite',
342: 'job',
343: 'birthday',

344: 'another',
345: 'lar',
346: 'camera',
347: 'dinner',
348: 'world',
349: 'account',
350: 'åf1000',
351: 'girl',
352: 'wanna',
353: 'dat',
354: 'boy',
355: 'åf1',
356: 'month',
357: 'forgot',
358: 'play',
359: 'den',
360: 'sweet',
361: 'awarded',
362: 'video',
363: 'wont',
364: 'bad',
365: 'speak',
366: 'shows',
367: 'rate',
368: 'ah',
369: 'probably',
370: 'guys',
371: 'check',
372: 'called',
373: 'enjoy',
374: 'liao',
375: 'landline',
376: 'face',
377: 'left',
378: 'bed',
379: 'watch',
380: 'remember',
381: 'apply',
382: 'll',
383: '2nd',
384: 'reach',
385: 'tv',
386: 'award',
387: 'cost',
388: 'room',
389: 'once',
390: 'nothing',
391: 'box',
392: 'shall',
393: 'xxx',

394: 'start',
395: 'between',
396: 'princess',
397: 'code',
398: 'does',
399: 'minutes',
400: 'two',
401: 'wife',
402: 'plan',
403: 'weekend',
404: 'maybe',
405: 'aight',
406: 'hour',
407: 'luv',
408: '10',
409: 'texts',
410: 'dunno',
411: 'bus',
412: 'everything',
413: 'most',
414: 'offer',
415: '000',
416: 'thanx',
417: 'sat',
418: '9',
419: '150ppm',
420: 'pa',
421: 'selected',
422: 'while',
423: 'missing',
424: 'those',
425: 'orange',
426: 'might',
427: 'pay',
428: 'says',
429: 'part',
430: 'actually',
431: 'kiss',
432: 'looking',
433: 'baby',
434: 'enough',
435: 'network',
436: 'jus',
437: 'hear',
438: 'fun',
439: 'put',
440: 'shopping',
441: 'asked',
442: 'ringtone',
443: 'update',

444: 'didn't',
445: 'bring',
446: 'early',
447: 'sexy',
448: 'dis',
449: 'mail',
450: 'easy',
451: 'years',
452: 'without',
453: 'po',
454: 'hav',
455: 'look',
456: 'plus',
457: 'must',
458: 'okay',
459: 'means',
460: 'mob',
461: "how's",
462: 'anyway',
463: 'bt',
464: 'watching',
465: 'evening',
466: 'thank',
467: 'fuck',
468: 'school',
469: 'tones',
470: 'xmas',
471: "he's",
472: 'entry',
473: 'abt',
474: 'collect',
475: 'join',
476: 'true',
477: 'important',
478: 'details',
479: 'afternoon',
480: 'made',
481: 'leh',
482: 'weekly',
483: '10p',
484: 'attempt',
485: 'collection',
486: 'colour',
487: 'since',
488: 'office',
489: 'tmr',
490: 'came',
491: 'bored',
492: 'wen',
493: 'dad',

494: 'until',
495: 'able',
496: 'price',
497: 'plz',
498: 'change',
499: 'pain',
500: 'goes',
501: 'wot',
502: 'dude',
503: 'these',
504: 'town',
505: 'run',
506: 'little',
507: '8',
508: 'wake',
509: 'm',
510: 'yesterday',
511: 'book',
512: 'calls',
513: 'messages',
514: 'worry',
515: 'though',
516: 'havent',
517: 'online',
518: 'test',
519: 'alright',
520: 'away',
521: 'wanted',
522: 'else',
523: 'national',
524: 'guy',
525: 'working',
526: 'game',
527: 'g',
528: 'saw',
529: 'smiling',
530: 'net',
531: 'wif',
532: 'valid',
533: 'making',
534: 'show',
535: 'sad',
536: 'words',
537: 'comes',
538: 'hair',
539: 'juz',
540: 'gift',
541: 'de',
542: 'missed',
543: 'trip',

544: 'driving',
545: 'both',
546: '500',
547: 'feeling',
548: 'stay',
549: 'drive',
550: 'lei',
551: 'club',
552: 'shop',
553: 'haf',
554: 'friendship',
555: 'huh',
556: 're',
557: 'order',
558: '8007',
559: 'double',
560: 'hours',
561: 'gr8',
562: 'hot',
563: 'å£5000',
564: 'bonus',
565: 'id',
566: 'todays',
567: 'wants',
568: 'sae',
569: "we're",
570: "''",
571: 'oso',
572: 'dreams',
573: 'decimal',
574: 'ring',
575: 'chikku',
576: 'times',
577: 'address',
578: 'eve',
579: 'http',
580: 'saying',
581: 'coz',
582: 'poly',
583: 'makes',
584: 'family',
585: 'top',
586: 'aft',
587: 'points',
588: 'tried',
589: 'pics',
590: 'food',
591: 'nite',
592: 'goin',
593: 'await',

594: 'pic',
595: 'âf2000',
596: 'date',
597: 'calling',
598: 'minute',
599: 'delivery',
600: 'mean',
601: 'old',
602: 'neva',
603: 'awesome',
604: 'news',
605: 'tomo',
606: 'leaving',
607: 'finished',
608: 'final',
609: 'âf100',
610: 'til',
611: 'hurt',
612: "there's",
613: 'busy',
614: 'knw',
615: 'question',
616: 'answer',
617: 'till',
618: 'rite',
619: 'story',
620: 'dating',
621: 'okie',
622: "haven't",
623: 'either',
624: 'everyone',
625: 'believe',
626: 'wid',
627: "she's",
628: 'simple',
629: 'touch',
630: 'full',
631: 'pounds',
632: 'post',
633: 'beautiful',
634: 'yours',
635: 'âf500',
636: 'sister',
637: 'private',
638: 'visit',
639: 'charge',
640: 'yourself',
641: 'noe',
642: "what's",
643: 'email',

644: 'happen',
645: 'sounds',
646: 'tot',
647: 'walk',
648: 'pm',
649: 'loving',
650: 'auction',
651: 'tho',
652: 'took',
653: 'movie',
654: 'forget',
655: 'xx',
656: 'available',
657: 'mobileupd8',
658: 'college',
659: 'drink',
660: 'fast',
661: 'type',
662: "won't",
663: 'sch',
664: 'choose',
665: 'lets',
666: 'together',
667: 'search',
668: 'statement',
669: 'light',
670: 'whats',
671: 'taking',
672: 'congrats',
673: 'hard',
674: 'smth',
675: 'w',
676: "we'll",
677: 'land',
678: 'needs',
679: 'second',
680: 'company',
681: 'pobox',
682: 'mine',
683: 'row',
684: 'sleeping',
685: '86688',
686: 'lose',
687: 'lots',
688: 'brother',
689: 'decided',
690: 'angry',
691: 'lesson',
692: 'gone',
693: 'ltd',

694: '100',
695: 'smoke',
696: 'found',
697: 'card',
698: 'wil',
699: '12hrs',
700: 'vouchers',
701: 'set',
702: 'identifier',
703: 'expires',
704: 'music',
705: '750',
706: '08000839402',
707: 'âf250',
708: "c's",
709: 'case',
710: 'takes',
711: 'o',
712: 'anyone',
713: 'b4',
714: 'happened',
715: 'yr',
716: 'fri',
717: 'unsubscribe',
718: 'head',
719: 'ard',
720: 'prob',
721: 'drop',
722: 'each',
723: 'fancy',
724: 'open',
725: 'close',
726: '\x89û',
727: 'cause',
728: 'pub',
729: 'etc',
730: 'least',
731: 'bout',
732: 'reason',
733: 'mum',
734: 'far',
735: 'saturday',
736: 'used',
737: 'thinking',
738: 'break',
739: 'âf3',
740: 'lucky',
741: 'quiz',
742: 'f',
743: 'weeks',

744: 'read',
745: 'goodmorning',
746: 'frnds',
747: 'credit',
748: 'party',
749: 'winner',
750: 'caller',
751: 'secret',
752: "doesn't",
753: 'treat',
754: 'services',
755: 'hows',
756: 'operator',
757: 'voucher',
758: 'rs',
759: '08000930705',
760: 'gas',
761: 'sun',
762: 'don',
763: 'msgs',
764: 'india',
765: '03',
766: 'opt',
767: 'sis',
768: 'tel',
769: 'seeing',
770: 'started',
771: 'hungry',
772: 'congratulations',
773: 'project',
774: 'hmm',
775: 'â€2',
776: 'download',
777: 'gd',
778: 'numbers',
779: 'â€350',
780: 'kind',
781: 'frnd',
782: 'happiness',
783: 'pretty',
784: 'dnt',
785: 'girls',
786: 'phones',
787: 'info',
788: 'darlin',
789: 'tc',
790: 'games',
791: 'chennai',
792: 'sunday',
793: 'march',

794: "you'll",
795: 'song',
796: 'area',
797: 'savamob',
798: 'luck',
799: 'friday',
800: 'ago',
801: 'point',
802: 'nope',
803: 'knew',
804: 'felt',
805: 'hold',
806: 'sending',
807: 'stupid',
808: 'yar',
809: 'mu',
810: 'boytoy',
811: 'course',
812: 'fone',
813: 'whole',
814: 'ass',
815: 'wit',
816: 'mates',
817: 'camcorder',
818: 'alone',
819: 'nt',
820: 'fucking',
821: 'complimentary',
822: 'ugh',
823: "you've",
824: 'side',
825: 'comp',
826: 'reading',
827: 'anytime',
828: 'park',
829: 'll',
830: 'motorola',
831: 'suite342',
832: '2lands',
833: 'understand',
834: 'wrong',
835: 'mom',
836: 'na',
837: 'ni8',
838: 'wkly',
839: 'father',
840: 'gal',
841: 'balance',
842: 'abiola',
843: 'hit',

844: 'oredi',
845: 'wine',
846: 'extra',
847: 'listen',
848: '20',
849: 'rental',
850: 'call2optout',
851: 'exam',
852: 'laptop',
853: 'their',
854: "'",
855: 'lovely',
856: 'bath',
857: 'sort',
858: 'worth',
859: 'offers',
860: 'jay',
861: 'log',
862: 'surprise',
863: 'hmmm',
864: 'sea',
865: 'frm',
866: 'ipod',
867: 'player',
868: 'pass',
869: '2003',
870: '800',
871: 'un',
872: 'couple',
873: 'truth',
874: '0800',
875: 'earlier',
876: 'reveal',
877: 'thinks',
878: 'shower',
879: 'hand',
880: '30',
881: 'crazy',
882: 'mate',
883: 'freemsg',
884: 'wonder',
885: 'snow',
886: 'difficult',
887: 'slow',
888: 'std',
889: 'carlos',
890: 'swing',
891: 'ten',
892: 'mah',
893: 'outside',

894: 'small',
895: 'finally',
896: 'content',
897: 'wonderful',
898: 'bslvyl',
899: 'bank',
900: 'valued',
901: 'p',
902: 'meant',
903: 'john',
904: 'mrng',
905: 'police',
906: 'currently',
907: 'wishing',
908: 'clean',
909: 'tickets',
910: 'th',
911: 'usf',
912: '12',
913: 'txts',
914: 'through',
915: 'warm',
916: 'reached',
917: 'tired',
918: 'myself',
919: 'askd',
920: 'darren',
921: 'redeemed',
922: 'feels',
923: 'rply',
924: 'store',
925: 'ts',
926: 'somebody',
927: 'rest',
928: 'months',
929: 'custcare',
930: 'welcome',
931: 'heard',
932: 'rakhesh',
933: 'parents',
934: 'kids',
935: 'rates',
936: 'hee',
937: 'usual',
938: 'link',
939: 'ex',
940: 'bcoz',
941: 'hoping',
942: 'checking',
943: 'entered',

944: 'sex',
945: 'talking',
946: 'own',
947: 'press',
948: 'age',
949: 'mr',
950: 'gym',
951: 'die',
952: 'ans',
953: 'via',
954: 'uncle',
955: 'otherwise',
956: 'direct',
957: "i'd",
958: 'eg',
959: "isn't",
960: 'lost',
961: 'kate',
962: 'match',
963: 'questions',
964: 'semester',
965: 'ha',
966: 'safe',
967: 'energy',
968: 'ac',
969: 'wana',
970: 'st',
971: 'england',
972: 'immediately',
973: 'loads',
974: 'blue',
975: 'empty',
976: "wasn't",
977: 'convey',
978: 'eh',
979: 'wow',
980: 'voice',
981: 'terms',
982: 'somewhere',
983: 'charity',
984: 'discount',
985: 'information',
986: 'fantastic',
987: 'worried',
988: 'study',
989: 'grins',
990: 'plans',
991: 'sell',
992: 'monday',
993: 'mobiles',

```

994: 'forever',
995: 'fact',
996: 'pray',
997: 'train',
998: 'against',
999: 'kinda',
1000: 'ending',
...}

```

Creating model and Adding the layers

```

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

```

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

```

Model: "sequential_2"

Layer (type)	Output Shape	Param #
=====		
embedding_1 (Embedding)	(None, 150, 32)	264032
lstm_1 (LSTM)	(None, 100)	53200
dropout_2 (Dropout)	(None, 100)	0
dense_2 (Dense)	(None, 20)	2020
dropout_3 (Dropout)	(None, 20)	0
dense_3 (Dense)	(None, 1)	21
=====		
Total params: 319,273		
Trainable params: 319,273		
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 [=====] - 12s 404ms/step - loss: 0.1012
- accuracy: 0.9799 - val_loss: 0.0636 - val_accuracy: 0.9831
Epoch 2/10
30/30 [=====] - 12s 392ms/step - loss: 0.0488
- accuracy: 0.9876 - val_loss: 0.0597 - val_accuracy: 0.9852

<keras.callbacks.History at 0x7f845ce77d90>

```

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 40ms/step - loss: 0.0646 -
accuracy: 0.9785

print('Test set\n Loss: {:.3f}\n Accuracy:
{:.3f}'.format(acc[0],acc[1]))

Test set
Loss: 0.065
Accuracy: 0.978

```