```
#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
                                                         v2 Unnamed: 2
/
         Go until jurong point, crazy.. Available only ...
                                                                   NaN
                             Ok lar... Joking wif u oni...
1
    ham
                                                                   NaN
   spam Free entry in 2 a wkly comp to win FA Cup fina...
2
                                                                   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
         NaN
                    NaN
1
2
         NaN
                    NaN
3
                    NaN
         NaN
         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
```

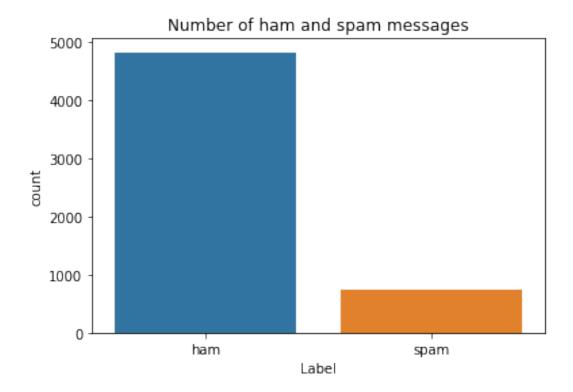
```
Data columns (total 2 columns):
     Column Non-Null Count
                             Dtype
 0
             5572 non-null
                             object
     v1
 1
     v2
             5572 non-null
                             object
dtypes: object(2)
memory usage: 87.2+ KB
from google.colab import drive
drive.mount('/content/drive')
Drive already mounted at /content/drive; to attempt to forcibly
remount, call drive.mount("/content/drive", force remount=True).
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')



#Process the Data

```
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)
\max \text{ 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([[
                Θ,
                     0, \ldots, 1, 28, 14
          0,
                     0, ..., 1, 64, 3],
0, ..., 140, 322, 195],
          0,
                0,
          0,
                0,
                     0, ..., 35, 7, 459],
                Θ,
          0,
                0,
                     0, ..., 128, 454, 981],
          0,
          0,
                Θ,
                     0, ..., 86, 245, 66]], dtype=int32)
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: '2',
 20: 'that',
21: 'are',
 22: 'now',
23: 'so',
 24: 'or',
 25: 'not',
 26: 'do',
```

```
27: 'but',
28: 'can',
29: 'be',
30: 'get',
31: 'at',
32: "i'm",
33: 'will',
34: 'ur',
35: 'with',
36: 'if',
37: 'just',
38: 'no',
39: 'this',
40: 'we',
41: 'gt',
42: 'Ĭt',
43: '4',
44: 'up',
45: 'when',
46: 'ok',
47: 'free',
48: 'out',
49: 'how',
50: 'go',
51: 'from',
52: 'all',
53: 'what',
54: 'like',
55: 'know',
56: 'then',
57: 'good',
58: 'was',
59: 'got',
60: 'come',
61: 'its',
62: 'am',
63: 'day',
64: 'love',
65: 'time',
66: 'only',
67: 'want',
68: 'there',
69: 'text',
70: 'he',
71: 'send',
72: 'as',
73: 'txt',
74: 'one',
75: 'by',
76: 'need',
```

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

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

```
177: 'sure',
178: 'babe',
179: 'cos',
180: 'won',
181: 'also',
182: 'let',
183: 'lol',
184: 'sent',
185: 'nokia',
186: 'wait',
187: 'would',
188: 'keep',
189: 'every',
190: 'uk',
191: '150p',
192: 'cant',
193: 'first',
194: 'care',
195: 'com',
196: 'before',
197: '5',
198: "i've",
199: 'us',
200: 'something',
201: 'urgent',
202: 'anything',
203: 'next',
204: 'late',
205: 's',
206: 'min',
207: 'contact',
208: 'b',
209: 'even',
210: 'someone',
211: 'pick',
212: 'buy',
213: 'again',
214: 'thing',
215: 'feel<sup>†</sup>,
216: 'tonight',
217: 'around',
218: 'place',
219: 'tone',
220: "can't",
221: 'over',
222: '50',
223: 'which',
224: 'sleep',
225: 'money',
226: 'things',
```

```
227: 'soon',
228: 'dun',
229: 'wan',
230: 'co',
231: 'went',
232: 'off',
233: 'ya',
234: 'were',
235: 'friends',
236: 'nice',
237: 'could',
238: 'his',
239: 'v',
240: 'always',
241: 'other',
242: 'down',
243: 'gonna',
244: '16',
245: '18',
246: 'service',
247: 'per',
248: 'many',
249: 'gud',
250: 'customer',
251: "that's",
252: 'fine',
253: 'x',
254: 'leave',
255: 'people',
256: 'wish',
257: 'chat',
258: 'sms',
259: 'haha',
260: 'year',
261: '6',
262: 'help',
263: 'heart',
264: 'special',
265: 'try',
266: 'friend',
267: "you're",
268: 'use',
269: 'told',
270: 'same',
271: 'mins',
272: 'thk',
273: 'yet',
274: 'guaranteed',
275: 'waiting',
276: 'coming',
```

```
277: 'live',
278: 'best',
279: 'cs',
280: 'lunch',
281: 'few',
     'may',
282:
283: 'thought',
284: 'never',
285: 'hello',
286: 'stuff',
287: 'ìï',
288: 'y',
289: 'trying',
290: '7',
291: 'draw',
292: 'name',
293: 'getting',
294: 'holiday',
295: 'god',
296: 'having',
297: 'finish',
298: 'days',
299: 'house',
300: 'thats',
301: 'cool',
302: 'man',
303: 'meeting',
304: "didn't",
305: 'done',
306: 'dat',
307: 'end',
308: 'class',
309: 'ready',
310: 'latest',
311: 'ill',
312: 'world',
313: 'line',
314: 'eat',
315: 'car',
316: 'wanna',
317: 'check',
318: 'half',
319: 'better',
320: 'lar',
321: 'person',
322: 'real',
323: 'because',
324: 'nothing',
325: 'guess',
326: 'long',
```

```
327: 'chance',
328: 'than',
329: 'å£1',
330: 'talk',
331: 'being',
332: 'yup',
333: 'enjoy',
334: 'play',
335: 'bit',
336: 'job',
337: 'shows',
338: 'smile',
339: 'mind',
340: 'another',
341: '1st',
342: 'box',
343: 'big',
344: 'aight',
345: 'problem',
346: 'receive',
347: 'guys',
348: 'dinner',
349: 'camera',
350: 'watching',
351: 'word',
352: 'å£1000',
353: 'cost',
354: 'room',
355: 'sweet',
356: 'start',
357: '150ppm',
358: 'jus',
359: 'boy',
360: 'awarded',
361: 'two',
362: 'account',
363: 'birthday',
364: 'wk',
365: 'lot',
366: 'ah',
367: 'apply',
368: 'month',
369: 'video',
370: 'sir',
371: 'maybe',
372: 'princess',
373: 'bt',
374: 'plan',
375: 'watch',
376: 'early',
```

```
377: 'might',
378: 'minutes',
379: 'called',
380: 'ever',
381: 'xxx',
382: '9',
383: 'into',
384: 'forgot',
385: 'hear',
386: 'liao',
387: 'offer',
388: 'quite',
389: 'luv',
390: 'girl',
391: 'm',
392: 'shit',
393: 'dunno',
394: 'bed',
395: 'dis',
396: 'landline',
397: "he's",
398: 'office',
399: 'between',
400: 'thank',
401: 'code',
402: 'does',
403: 'weekend',
404: 'part',
405: 'shall',
406: 'yo',
407: 'pa',
408: '10',
409: 'hour',
410: 'fun',
411: 'without',
412: 'easy',
413: 'asked',
414: 'rate',
415: 'kiss',
416: 'orange',
417: 'look',
418: 'leh',
419: 'selected',
420: 'baby',
421: 'little',
422: 'face',
423: 'wont',
424: 'left',
425: 'working',
426: 'll',
```

```
427: 'details',
428: 'missing',
429: 'network',
430: 'bad',
431: '000',
432: 'didnt',
433: 'town',
434: 'probably',
435: 'speak',
436: 'ringtone',
437: 'reach',
438: 'den',
439: 'po',
440: 'fuck',
441: 'put',
442: 'entry',
443: 'once',
444: 'xmas',
445: 'sexy',
446: 'most',
447: 'since',
448: 'sat',
449: 'thanx',
450: 'hav',
451: 'actually',
452: 'looking',
453: 'shopping',
454: 'until',
455: 'tv',
456: 'means',
457: 'though',
458: 'made',
459: 'dad',
460: 'pay',
461: 'update',
462: 'collect',
463: 'join',
464: 'texts',
465: 'award',
466: 'bus',
467: 'everything',
468: 'enough',
469: 'wanted',
470: 'evening',
471: "how's",
472: 'anyway',
473: 'important',
474: "there's",
475: 'plus',
476: 'those',
```

```
477: 'afternoon',
478: '8',
479: 'g',
480: 'remember',
481: 'å£100',
482: 'school',
483: 'gift',
484: 'de',
485: 'pain',
486: 'able',
487: 'says',
488: '2nd',
489: 'came',
490: 'tones',
491: 'wake',
492: 'tmr',
493: 'till',
494: 'double',
495: 'wif',
496: 'music',
497: 'abt',
498: 'bring',
499: 'while',
500: 'away',
501: 'true',
502: 'must',
503: 'wife',
504: 'wot',
505: 'saw',
506: 'else',
507: 'juz',
508: 'book',
509: 'alright',
510: 'worry',
511: 'price',
512: 'either',
513: 'net',
514: 'colour',
515: 'run',
516: 'collection',
517: 'online',
518: 'weekly',
519: 'feeling',
520: '500',
521: 'oso',
522: 'making',
523: 'question',
524: 'mail',
525: 'okay',
526: 'coz',
```

```
527: 'til',
528: 'show',
529: 'friendship',
530: 'game',
531: 'these'
532: "what's",
533: 'decimal',
534: 'calls',
535: 'messages',
536: 'plz',
537: 'family',
538: 'yourself',
539: 'attempt',
540: 'havent',
541: 'full',
542: '10p',
543: 'years',
544: 'guy',
545: 'hurt',
546: 'sad',
547: 'vouchers',
548: 'answer',
549: 'words',
550: 'mob',
551: 'hot',
552: 'yesterday',
553: 'id',
554: 'dude',
555: 'goes',
556: 'makes',
557: 'driving',
558: 'points',
559: 'nite',
560: 'valid',
561: "haven't",
562: 'shop',
563: 'set',
564: 'wants',
565: 'times',
566: "won't",
567: 'http',
568: 'date',
569: 'comes',
570: 'both',
571: 'test',
572: 'busy',
573: 'eve',
574: 'happen',
575: 'games',
576: 'charge',
```

```
577: 'å£5000',
578: 'close',
579: 'trip',
580: 'todays',
581: 'believe',
582: 'pic',
583: 'movie',
584: 'bored',
585: 'wen',
586: 're',
587: 'cause',
588: 'ard',
589: 'neva',
590: 'sae',
591: 'å£500',
592: 'yours',
593: 'ĺose',
594: 'lei',
595: 'order',
596: 'hours',
597: 'pounds',
598: 'story',
599: '8007',
600: 'national',
601: 'top',
602: 'hair',
603: 'noe',
604: 'stay',
605: 'tot',
606: 'food',
607: 'lets',
608: 'private',
609: 'whats',
610: 'w',
611: 'missed',
612: 'email',
613: 'change',
614: 'ring',
615: 'post',
616: 'saying',
617: 'card',
618: 'available',
619: 'old',
620: 'pm',
621: 'leaving',
622: 'haf',
623: 'dreams',
624: 'news',
625: 'forget',
626: 'delivery',
```

```
627: 'second',
628: 'started',
629: 'await',
630: 'beautiful',
631: 'finished',
632: 'aft',
633: 'bonus',
634: 'together',
635: 'rite',
636: 'walk',
637: 'congrats',
638: 'mean',
639: '11',
640: 'worth',
641: 'huh',
642: 'taking',
643: 'expires',
644: 'drink',
645: 'everyone',
646: 'lucky',
647: 'lots',
648: 'pub',
649: 'auction',
650: 'å£2000',
651: 'pics',
652: 'decided',
653: 'final',
654: 'alone',
655: 'takes',
656: 'tried',
657: 'chikku',
658: "she's",
659: 'tho',
660: 'drop',
661: "c's",
662: 'address',
663: 'simple',
664: 'tomo',
665: '86688'
666: 'pobox',
667: 'ltd',
668: 'winner',
669: 'don',
670: 'congratulations',
671: "you've",
672: 'touch',
673: 'calling',
674: 'choose',
675: 'statement'
676: 'identifier',
```

```
677: 'company',
678: 'drive',
679: 'visit',
680: 'found',
681: 'row',
682: 'hows',
683: 'goin',
684: 'prob',
685: 'thinking',
686: 'land',
687: 'services',
688: 'okie',
689: 'poly',
690: 'type',
691: 'smth',
692: '08000930705',
693: 'sch',
694: 'kind',
695: 'club',
696: '750',
697: 'fast',
698: 'wid',
699: 'minute',
700: 'whatever',
701: "we're",
702: 'qd',
703: 'smoke',
704: "''",
705: 'head',
706: 'knw',
707: 'far',
708: 'mobileupd8'
709: '08000839402',
710: '100',
711: 'hand',
712: "we'll",
713: 'oredi',
714: 'open',
715: 'happened',
716: 'lesson',
717: 'darlin',
718: 'hard',
719: 'å£3',
720: 'camcorder',
721: 'operator',
722: '30',
723: 'msgs',
724: 'friday',
725: 'gone',
726: 'anytime',
```

```
727: 'voucher',
728: 'break',
729: 'awesome',
730: 'mine',
731: 'wil',
732: 'mum',
733: 'search',
734: 'took',
735: 'sleeping',
736: 'savamob',
737: 'offers',
738: 'frnd',
739: "doesn't",
740: 'crazy',
741: 'goodmorning',
742: 'mates',
743: 'fr',
744: 'loving',
745: 'pretty',
746: 'wkly',
747: 'support',
748: 'whole',
749: 'gr8',
750: 'press',
751: 'party',
752: 'happiness',
753: 'sounds',
754: 'india',
755: 'quiz',
756: 'sun',
757: 'xx',
758: '12hrs',
759: 'unsubscribe',
760: 'gas',
761: 'mate',
762: "'",
763: 'brother',
764: 'sex',
765: 'bout'
766: 'credit',
767: 'opt',
768: 'o',
769: 'fone',
770: 'dating',
771: 'nt',
772: 'angry',
773: 'smiling',
774: 'each',
775: 'ni8',
776: 'boytoy',
```

```
777: 'un',
778: 'wrong',
779: 'saturday',
780: 'case',
781: 'mr',
782: 'used',
783: 'sister',
784: 'tc',
785: 'cut',
786: 'wit',
787: '\x89û',
788: 'lovely',
789: 'info',
790: 'ends',
791: 'phones',
792: '87066',
793: 'ans',
794: 'treat',
795: 'hee',
796: 'mah',
797: 'carlos',
798: 'gal',
799: 'knew',
800: 'telling',
801: 'dnt',
802: 'least',
803: 'project',
804: 'light',
805: 'lost',
806: 'welcome',
807: 'anyone',
808: 'suite342',
809: '2lands',
810: 'cum',
811: 'currently',
812: 'parents',
813: 'etc',
814: 'almost',
815: 'fucking',
816: 'fancy',
817: 'gym',
818: 'couple',
819: 'motorola',
820: 'chennai',
821: 'sis',
822: 'hold',
823: 'å£2',
824: '03',
825: 'girls',
826: 'mobiles',
```

```
827: 'understand',
828: 'tel',
829: 'mom',
830: 'unlimited',
831: 'mu',
832: 'wonderful',
833: 'luck',
834: 'through',
835: 'park',
836: 'b4',
837: 'ten',
838: 'bslvyl',
839: 'ago',
840: '2003',
841: '800',
842: 'redeemed',
843: '04',
844: 'snow',
845: 'hit',
846: 'outside',
847: 'fri',
848: 'secret',
849: 'weeks',
850: 'player',
851: 'f',
852: 'å£250',
853: 'frnds',
854: 'crave',
855: "you'll",
856: 'hmmm',
857: 'meant',
858: 'nope',
859: 'remove',
860: 'å£350',
861: 'talking',
862: 'read',
863: 'christmas',
864: 'college',
865: 'numbers',
866: 'difficult',
867: 'listen',
868: 'finally',
869: 'abiola',
870: 'course',
871: 'seeing',
872: 'st',
873: 'tired',
874: 'age',
875: 'caller',
876: 'information',
```

```
877: 'log',
878: 'mrng',
879: 'feels',
880: 'earlier',
881: 'th',
882: 'loads',
883: 'dogging',
884: 'yr',
885: 'rest',
886: 'blue',
887: 'safe',
888: 'within',
889: '20',
890: 'ac',
891: 'laptop',
892: 'slowly',
893: 'clean',
894: 'thinks',
895: 'direct',
896: 'father',
897: 'their',
898: 'sea',
899: 'yar',
900: 'frm',
901: 'via',
902: 'semester',
903: 'askd',
904: 'loved',
905: 'comin',
906: 'die',
907: 'comp',
908: 'ipod',
909: 'supposed',
910: 'gn',
911: 'txting',
912: 'invited',
913: 'film',
914: 'slow',
915: 'knows',
916: 'write',
917: 'felt',
918: 'point',
919: 'met',
920: 'std',
921: '0800',
922: 'rental',
923: 'hmm',
924: 'gets',
925: 'figure',
926: 'rs',
```

```
927: 'mayb',
928: 'orchard',
929: 'balance',
930: 'freemsg',
931: 'checking',
932: 'hungry',
933: 'reward',
934: 'eh',
935: 'england',
936: 'call2optout',
937: 'callertune',
938: 'voice',
939: 'reason',
940: 'exam',
941: 'area',
942: 'jay',
943: 'seen',
944: 'bathe',
945: 'discount',
946: 'ass',
947: 'bank',
948: 'mm',
949: 'surprise',
950: 'lovable',
951: 'song',
952: '00',
953: 'truth',
954: 'red',
955: 'shower',
956: 'computer',
957: 'correct',
958: 'small',
959: 'blood',
960: 'monday',
961: 'charged',
962: 'confirm',
963: 'complimentary',
964: 'hospital',
965: 'sort',
966: 'ge',
967: 'pass',
968: 'stupid',
969: "wasn't",
970: 'picking',
971: 'na',
972: 'march',
973: 'woke',
974: 'content',
975: 'rock',
976: 'empty',
```

```
977: 'rply',
 978: 'paper',
 979: 'gotta',
 980: 'heard',
 981: 'sunday',
 982: 'sound',
 983: 'ntt',
 984: 'tonite',
 985: 'water',
 986: 'somewhere',
 987: 'dead',
 988: 'lessons',
 989: 'ive',
 990: 'match',
 991:
      'months',
 992: 'cd',
 993: 'enter',
 994: 'doin',
 995: 'valentines',
      'oops',
 996:
 997: 'different',
 998: 'ugh',
999: 'download',
 1000: 'al',
 . . . }
#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 1"
                              Output Shape
Layer (type)
                                                          Param #
 embedding 1 (Embedding)
                              (None, 150, 32)
                                                          262240
 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: 317,481
Trainable params: 317,481
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
- accuracy: 0.8593 - val loss: 0.2674 - val accuracy: 0.8650
Epoch 2/10
- accuracy: 0.9430 - val loss: 0.1178 - val accuracy: 0.9747
<keras.callbacks.History at 0x7f35ed115790>
#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)
accuracy: 0.9725
print('Test set\n Loss: {:0.3f}\n Accuracy:
{:0.3f}'.format(acc[0],acc[1]))
Test set
Loss: 0.114
Accuracy: 0.972
```