news

January 12, 2025

0.1

: , LSTM, GRU.

```python
[1]: %tensorflow_version 2.x
     from tensorflow.keras.models import Sequential
     from tensorflow.keras.layers import Dense, Embedding, MaxPooling1D, Conv1D,
      ↪GlobalMaxPooling1D, Dropout, LSTM, GRU
     from tensorflow.keras import utils
     from tensorflow.keras.preprocessing.sequence import pad_sequences
     from tensorflow.keras.preprocessing.text import Tokenizer
     from tensorflow.keras.callbacks import ModelCheckpoint
     from tensorflow.keras import utils
     import pandas as pd
     import numpy as np
     import matplotlib.pyplot as plt
     %matplotlib inline
```

Colab only includes TensorFlow 2.x; %tensorflow_version has no effect.

```python
[2]: #
     num_words = 10000
     #
     max_news_len = 30
     #
     nb_classes = 4
```

0.2

```python
[3]: !wget https://github.com/mhjabreel/CharCnn_Keras/raw/master/data/ag_news_csv/
      ↪train.csv -O train.csv
```

```
--2025-01-12 10:45:36--
https://github.com/mhjabreel/CharCnn_Keras/raw/master/data/ag_news_csv/train.csv
Resolving github.com (github.com)… 140.82.114.4
Connecting to github.com (github.com)|140.82.114.4|:443… connected.
HTTP request sent, awaiting response… 302 Found
```

```
Location: https://raw.githubusercontent.com/mhjabreel/CharCnn_Keras/master/data/
ag_news_csv/train.csv [following]
--2025-01-12 10:45:36--  https://raw.githubusercontent.com/mhjabreel/CharCnn_Ker
as/master/data/ag_news_csv/train.csv
Resolving raw.githubusercontent.com (raw.githubusercontent.com)…
185.199.108.133, 185.199.109.133, 185.199.110.133, …
Connecting to raw.githubusercontent.com
(raw.githubusercontent.com)|185.199.108.133|:443… connected.
HTTP request sent, awaiting response… 200 OK
Length: 29470338 (28M) [text/plain]
Saving to: 'train.csv'

train.csv             100%[===================>]  28.10M  --.-KB/s    in 0.1s

2025-01-12 10:45:37 (246 MB/s) - 'train.csv' saved [29470338/29470338]
```

[4]: `!wget https://github.com/mhjabreel/CharCnn_Keras/raw/master/data/ag_news_csv/`
`↪test.csv -O test.csv`

```
--2025-01-12 10:45:37--
https://github.com/mhjabreel/CharCnn_Keras/raw/master/data/ag_news_csv/test.csv
Resolving github.com (github.com)… 140.82.116.3
Connecting to github.com (github.com)|140.82.116.3|:443… connected.
HTTP request sent, awaiting response… 302 Found
Location: https://raw.githubusercontent.com/mhjabreel/CharCnn_Keras/master/data/
ag_news_csv/test.csv [following]
--2025-01-12 10:45:38--  https://raw.githubusercontent.com/mhjabreel/CharCnn_Ker
as/master/data/ag_news_csv/test.csv
Resolving raw.githubusercontent.com (raw.githubusercontent.com)…
185.199.108.133, 185.199.109.133, 185.199.110.133, …
Connecting to raw.githubusercontent.com
(raw.githubusercontent.com)|185.199.108.133|:443… connected.
HTTP request sent, awaiting response… 200 OK
Length: 1857427 (1.8M) [text/plain]
Saving to: 'test.csv'

test.csv              100%[===================>]   1.77M  --.-KB/s    in 0.05s

2025-01-12 10:45:38 (34.0 MB/s) - 'test.csv' saved [1857427/1857427]
```

[5]: `!wget https://raw.githubusercontent.com/mhjabreel/CharCnn_Keras/master/data/`
`↪ag_news_csv/classes.txt -O classes.txt`

```
--2025-01-12 10:45:38--  https://raw.githubusercontent.com/mhjabreel/CharCnn_Ker
as/master/data/ag_news_csv/classes.txt
Resolving raw.githubusercontent.com (raw.githubusercontent.com)…
185.199.108.133, 185.199.109.133, 185.199.111.133, …
Connecting to raw.githubusercontent.com
(raw.githubusercontent.com)|185.199.108.133|:443… connected.
HTTP request sent, awaiting response… 200 OK
Length: 31 [text/plain]
Saving to: 'classes.txt'

classes.txt         100%[===================>]      31  --.-KB/s    in 0s

2025-01-12 10:45:38 (787 KB/s) - 'classes.txt' saved [31/31]
```

0.3

[6]:
```
!ls
```

```
classes.txt  sample_data  test.csv  train.csv
```

[7]:
```
!cat classes.txt
```

```
World
Sports
Business
Sci/Tech
```

[8]:
```
!head train.csv
```

```
"3","Wall St. Bears Claw Back Into the Black (Reuters)","Reuters - Short-
sellers, Wall Street's dwindling\band of ultra-cynics, are seeing green again."
"3","Carlyle Looks Toward Commercial Aerospace (Reuters)","Reuters - Private
investment firm Carlyle Group,\which has a reputation for making well-timed and
occasionally\controversial plays in the defense industry, has quietly placed\its
bets on another part of the market."
"3","Oil and Economy Cloud Stocks' Outlook (Reuters)","Reuters - Soaring crude
prices plus worries\about the economy and the outlook for earnings are expected
to\hang over the stock market next week during the depth of the\summer
doldrums."
"3","Iraq Halts Oil Exports from Main Southern Pipeline (Reuters)","Reuters -
Authorities have halted oil export\flows from the main pipeline in southern Iraq
after\intelligence showed a rebel militia could strike\infrastructure, an oil
official said on Saturday."
"3","Oil prices soar to all-time record, posing new menace to US economy
(AFP)","AFP - Tearaway world oil prices, toppling records and straining wallets,
present a new economic menace barely three months before the US presidential
elections."
```

"3","Stocks End Up, But Near Year Lows (Reuters)","Reuters - Stocks ended slightly higher on Friday\but stayed near lows for the year as oil prices surged past #36;46\a barrel, offsetting a positive outlook from computer maker\Dell Inc. (DELL.O)"
"3","Money Funds Fell in Latest Week (AP)","AP - Assets of the nation's retail money market mutual funds fell by #36;1.17 billion in the latest week to #36;849.98 trillion, the Investment Company Institute said Thursday."
"3","Fed minutes show dissent over inflation (USATODAY.com)","USATODAY.com - Retail sales bounced back a bit in July, and new claims for jobless benefits fell last week, the government said Thursday, indicating the economy is improving from a midsummer slump."
"3","Safety Net (Forbes.com)","Forbes.com - After earning a PH.D. in Sociology, Danny Bazil Riley started to work as the general manager at a commercial real estate firm at an annual base salary of #36;70,000. Soon after, a financial planner stopped by his desk to drop off brochures about insurance benefits available through his employer. But, at 32, ""buying insurance was the furthest thing from my mind,"" says Riley."
"3","Wall St. Bears Claw Back Into the Black"," NEW YORK (Reuters) - Short-sellers, Wall Street's dwindling band of ultra-cynics, are seeing green again."

[9]: `!head test.csv`

"3","Fears for T N pension after talks","Unions representing workers at Turner Newall say they are 'disappointed' after talks with stricken parent firm Federal Mogul."
"4","The Race is On: Second Private Team Sets Launch Date for Human Spaceflight (SPACE.com)","SPACE.com - TORONTO, Canada -- A second\team of rocketeers competing for the #36;10 million Ansari X Prize, a contest for\privately funded suborbital space flight, has officially announced the first\launch date for its manned rocket."
"4","Ky. Company Wins Grant to Study Peptides (AP)","AP - A company founded by a chemistry researcher at the University of Louisville won a grant to develop a method of producing better peptides, which are short chains of amino acids, the building blocks of proteins."
"4","Prediction Unit Helps Forecast Wildfires (AP)","AP - It's barely dawn when Mike Fitzpatrick starts his shift with a blur of colorful maps, figures and endless charts, but already he knows what the day will bring. Lightning will strike in places he expects. Winds will pick up, moist places will dry and flames will roar."
"4","Calif. Aims to Limit Farm-Related Smog (AP)","AP - Southern California's smog-fighting agency went after emissions of the bovine variety Friday, adopting the nation's first rules to reduce air pollution from dairy cow manure."
"4","Open Letter Against British Copyright Indoctrination in Schools","The British Department for Education and Skills (DfES) recently launched a ""Music Manifesto"" campaign, with the ostensible intention of educating the next generation of British musicians. Unfortunately, they also teamed up with the music industry (EMI, and various artists) to make this popular. EMI has apparently negotiated their end well, so that children in our schools will now

be indoctrinated about the illegality of downloading music.The ignorance and audacity of this got to me a little, so I wrote an open letter to the DfES about it. Unfortunately, it's pedantic, as I suppose you have to be when writing to goverment representatives. But I hope you find it useful, and perhaps feel inspired to do something similar, if or when the same thing has happened in your area."
"4","Loosing the War on Terrorism","\\""Sven Jaschan, self-confessed author of the Netsky and Sasser viruses, is\responsible for 70 percent of virus infections in 2004, according to a six-month\virus roundup published Wednesday by antivirus company Sophos.""\\""The 18-year-old Jaschan was taken into custody in Germany in May by police who\said he had admitted programming both the Netsky and Sasser worms, something\experts at Microsoft confirmed. (A Microsoft antivirus reward program led to the\teenager's arrest.) During the five months preceding Jaschan's capture, there\were at least 25 variants of Netsky and one of the port-scanning network worm\Sasser.""\\""Graham Cluley, senior technology consultant at Sophos, said it was staggeri …\\"
"4","FOAFKey: FOAF, PGP, Key Distribution, and Bloom Filters","\\FOAF/LOAF and bloom filters have a lot of interesting properties for social\network and whitelist distribution.\\I think we can go one level higher though and include GPG/OpenPGP key\fingerpring distribution in the FOAF file for simple web-of-trust based key\distribution.\\What if we used FOAF and included the PGP key fingerprint(s) for identities?\This could mean a lot. You include the PGP key fingerprints within the FOAF\file of your direct friends and then include a bloom filter of the PGP key\fingerprints of your entire whitelist (the source FOAF file would of course need\to be encrypted).\\Your whitelist would be populated from the social network as your client\discovered new identit …\\"
"4","E-mail scam targets police chief","Wiltshire Police warns about ""phishing"" after its fraud squad chief was targeted."
"4","Card fraud unit nets 36,000 cards","In its first two years, the UK's dedicated card fraud unit, has recovered 36,000 stolen cards and 171 arrests - and estimates it saved 65m."

```
[10]: !wc -l train.csv
      !wc -l test.csv
```

```
120000 train.csv
7600 test.csv
```

0.4

```
[11]: train = pd.read_csv('train.csv',
                          header=None,
                          names=['class', 'title', 'text'])
```

```
[12]: train
```

```
[12]:         class                                                     title  \
       0          3  Wall St. Bears Claw Back Into the Black (Reuters)
       1          3  Carlyle Looks Toward Commercial Aerospace (Reu…
       2          3     Oil and Economy Cloud Stocks' Outlook (Reuters)
       3          3  Iraq Halts Oil Exports from Main Southern Pipe…
       4          3  Oil prices soar to all-time record, posing new…
       …        …                                                     …
       119995     1  Pakistan's Musharraf Says Won't Quit as Army C…
       119996     2                      Renteria signing a top-shelf deal
       119997     2                           Saban not going to Dolphins yet
       119998     2                                     Today's NFL games
       119999     2                           Nets get Carter from Raptors

                                                           text
       0         Reuters - Short-sellers, Wall Street's dwindli…
       1         Reuters - Private investment firm Carlyle Grou…
       2         Reuters - Soaring crude prices plus worries\ab…
       3         Reuters - Authorities have halted oil export\f…
       4         AFP - Tearaway world oil prices, toppling reco…
       …                                                       …
       119995    KARACHI (Reuters) - Pakistani President Perve…
       119996   Red Sox general manager Theo Epstein acknowled…
       119997   The Miami Dolphins will put their courtship of…
       119998   PITTSBURGH at NY GIANTS Time: 1:30 p.m. Line: …
       119999   INDIANAPOLIS -- All-Star Vince Carter was trad…

       [120000 rows x 3 columns]
```

```
[13]: news = train['text']
```

```
[14]: news[:5]
```

```
[14]: 0    Reuters - Short-sellers, Wall Street's dwindli…
      1    Reuters - Private investment firm Carlyle Grou…
      2    Reuters - Soaring crude prices plus worries\ab…
      3    Reuters - Authorities have halted oil export\f…
      4    AFP - Tearaway world oil prices, toppling reco…
      Name: text, dtype: object
```

```
[15]: y_train = utils.to_categorical(train['class'] - 1, nb_classes)
```

```
[16]: y_train
```

```
[16]: array([[0., 0., 1., 0.],
              [0., 0., 1., 0.],
              [0., 0., 1., 0.],
              …,
              [0., 1., 0., 0.],
              [0., 1., 0., 0.],
              [0., 1., 0., 0.]])
```

0.5

```
[17]: news[:5]
```

```
[17]: 0    Reuters - Short-sellers, Wall Street's dwindli…
      1    Reuters - Private investment firm Carlyle Grou…
      2    Reuters - Soaring crude prices plus worries\ab…
      3    Reuters - Authorities have halted oil export\f…
      4    AFP - Tearaway world oil prices, toppling reco…
      Name: text, dtype: object
```

Keras

```
[18]: tokenizer = Tokenizer(num_words=num_words)
```

```
[19]: tokenizer.fit_on_texts(news)
```

```
[20]: tokenizer.word_index
```

```
[20]: {'the': 1,
       'a': 2,
       'to': 3,
       'of': 4,
       'in': 5,
       'and': 6,
       'on': 7,
       'for': 8,
       '39': 9,
       's': 10,
       'that': 11,
       'with': 12,
       'as': 13,
       'its': 14,
       'at': 15,
       'said': 16,
       'is': 17,
```

```
'by': 18,
'it': 19,
'has': 20,
'new': 21,
'an': 22,
'from': 23,
'reuters': 24,
'his': 25,
'will': 26,
'was': 27,
'after': 28,
'have': 29,
'be': 30,
'their': 31,
'two': 32,
'are': 33,
'us': 34,
'over': 35,
'quot': 36,
'year': 37,
'first': 38,
'ap': 39,
'he': 40,
'but': 41,
'gt': 42,
'lt': 43,
'this': 44,
'more': 45,
'monday': 46,
'wednesday': 47,
'one': 48,
'tuesday': 49,
'up': 50,
'thursday': 51,
'company': 52,
'inc': 53,
'friday': 54,
'world': 55,
'than': 56,
'u': 57,
'1': 58,
'last': 59,
'they': 60,
'york': 61,
'yesterday': 62,
'against': 63,
'about': 64,
```

```
'who': 65,
'not': 66,
'were': 67,
'into': 68,
'out': 69,
'three': 70,
'been': 71,
'president': 72,
'2': 73,
'had': 74,
'million': 75,
'corp': 76,
'oil': 77,
'when': 78,
'week': 79,
'time': 80,
'would': 81,
'united': 82,
'sunday': 83,
'which': 84,
'game': 85,
'people': 86,
'today': 87,
'government': 88,
'years': 89,
'could': 90,
'no': 91,
'second': 92,
'group': 93,
'percent': 94,
'n': 95,
'com': 96,
'saturday': 97,
'software': 98,
'next': 99,
'all': 100,
'third': 101,
'season': 102,
'night': 103,
'or': 104,
'prices': 105,
'iraq': 106,
'security': 107,
'day': 108,
'3': 109,
'fullquote': 110,
'quarter': 111,
```

```
'off': 112,
'stocks': 113,
'6': 114,
'microsoft': 115,
'minister': 116,
'announced': 117,
'internet': 118,
'team': 119,
'some': 120,
'four': 121,
'back': 122,
'state': 123,
'international': 124,
'high': 125,
'washington': 126,
'2004': 127,
'billion': 128,
'may': 129,
'most': 130,
'market': 131,
'news': 132,
'former': 133,
'10': 134,
'officials': 135,
'top': 136,
'can': 137,
'4': 138,
'business': 139,
'says': 140,
'other': 141,
'win': 142,
'states': 143,
'if': 144,
'month': 145,
'5': 146,
'victory': 147,
'city': 148,
'record': 149,
'end': 150,
'before': 151,
'european': 152,
'open': 153,
'largest': 154,
'technology': 155,
'just': 156,
'co': 157,
'service': 158,
```

```
'reported': 159,
't': 160,
'american': 161,
'league': 162,
'7': 163,
'sales': 164,
'afp': 165,
'computer': 166,
'home': 167,
'down': 168,
'federal': 169,
'000': 170,
'you': 171,
'five': 172,
'killed': 173,
'national': 174,
'according': 175,
'prime': 176,
'what': 177,
'expected': 178,
'made': 179,
'plans': 180,
'research': 181,
'while': 182,
'0': 183,
'major': 184,
'ticker': 185,
'network': 186,
'target': 187,
'now': 188,
'during': 189,
'least': 190,
'court': 191,
'online': 192,
'country': 193,
'between': 194,
'chief': 195,
'maker': 196,
'http': 197,
'companies': 198,
'london': 199,
'under': 200,
'them': 201,
'long': 202,
'san': 203,
'www': 204,
'british': 205,
```

```
'china': 206,
'set': 207,
'lead': 208,
'cup': 209,
'deal': 210,
'bush': 211,
'final': 212,
'there': 213,
'series': 214,
'href': 215,
'another': 216,
'search': 217,
'make': 218,
'based': 219,
'south': 220,
'bank': 221,
'report': 222,
'since': 223,
'her': 224,
'take': 225,
'police': 226,
'won': 227,
'old': 228,
'investor': 229,
'space': 230,
'giant': 231,
'games': 232,
'coach': 233,
'help': 234,
'john': 235,
'being': 236,
'industry': 237,
'through': 238,
'sports': 239,
'left': 240,
'music': 241,
'shares': 242,
'election': 243,
'leader': 244,
'run': 245,
'him': 246,
'services': 247,
'way': 248,
'web': 249,
'agreed': 250,
'so': 251,
'say': 252,
```

```
'only': 253,
'iraqi': 254,
'red': 255,
'because': 256,
'hit': 257,
'early': 258,
'system': 259,
'mobile': 260,
'military': 261,
'six': 262,
'aspx': 263,
'quickinfo': 264,
'o': 265,
'months': 266,
'profit': 267,
'like': 268,
'baghdad': 269,
'general': 270,
'i': 271,
'days': 272,
'north': 273,
'al': 274,
'get': 275,
'rose': 276,
'p': 277,
'war': 278,
'including': 279,
'many': 280,
'union': 281,
'still': 282,
'biggest': 283,
'ago': 284,
'even': 285,
'executive': 286,
'half': 287,
'big': 288,
'talks': 289,
'strong': 290,
'8': 291,
'official': 292,
'palestinian': 293,
'trade': 294,
'india': 295,
'plan': 296,
'wireless': 297,
'bid': 298,
'players': 299,
```

```
'data': 300,
'round': 301,
'latest': 302,
'phone': 303,
'held': 304,
'higher': 305,
'olympic': 306,
'much': 307,
'start': 308,
'released': 309,
'points': 310,
'stock': 311,
'growth': 312,
'move': 313,
'part': 314,
'nuclear': 315,
'athens': 316,
'earnings': 317,
'google': 318,
'test': 319,
'boston': 320,
'11': 321,
'west': 322,
'20': 323,
'how': 324,
'where': 325,
'fourth': 326,
'dollar': 327,
'called': 328,
'global': 329,
'head': 330,
'investors': 331,
'play': 332,
'australia': 333,
'public': 334,
'face': 335,
'weeks': 336,
'england': 337,
'israeli': 338,
'users': 339,
'presidential': 340,
'nearly': 341,
'seven': 342,
'economic': 343,
'football': 344,
'found': 345,
'windows': 346,
```

```
'japan': 347,
'air': 348,
'nations': 349,
'financial': 350,
'despite': 351,
'ahead': 352,
'francisco': 353,
'cut': 354,
'ibm': 355,
'man': 356,
'free': 357,
'gold': 358,
'took': 359,
'championship': 360,
'around': 361,
'amp': 362,
'foreign': 363,
'fell': 364,
'systems': 365,
'go': 366,
'12': 367,
'also': 368,
'9': 369,
'street': 370,
'firm': 371,
'russian': 372,
'number': 373,
'should': 374,
'near': 375,
'loss': 376,
'any': 377,
'leading': 378,
'used': 379,
'right': 380,
'forces': 381,
'past': 382,
'reports': 383,
'use': 384,
'following': 385,
'do': 386,
'buy': 387,
'troops': 388,
'30': 389,
'quote': 390,
'work': 391,
'key': 392,
'attack': 393,
```

```
'sox': 394,
'pay': 395,
'drug': 396,
'economy': 397,
'chicago': 398,
'release': 399,
'agency': 400,
'led': 401,
'36': 402,
'share': 403,
'car': 404,
'both': 405,
'15': 406,
'per': 407,
'profile': 408,
'version': 409,
'october': 410,
'september': 411,
'press': 412,
'late': 413,
'commission': 414,
'gaza': 415,
'beat': 416,
'video': 417,
'media': 418,
'killing': 419,
'your': 420,
'2005': 421,
'price': 422,
'player': 423,
'best': 424,
'rival': 425,
'put': 426,
'several': 427,
'uk': 428,
'apple': 429,
'america': 430,
'well': 431,
'again': 432,
'wall': 433,
'power': 434,
'capital': 435,
'products': 436,
'here': 437,
'tokyo': 438,
'source': 439,
'contract': 440,
```

```
'offer': 441,
'close': 442,
'e': 443,
'agreement': 444,
'region': 445,
'such': 446,
'australian': 447,
'recent': 448,
'workers': 449,
'french': 450,
'support': 451,
'might': 452,
'making': 453,
'program': 454,
'conference': 455,
'demand': 456,
'leaders': 457,
'scientists': 458,
'whether': 459,
'17': 460,
'title': 461,
'customers': 462,
'crude': 463,
'digital': 464,
'attacks': 465,
'scored': 466,
'un': 467,
'champion': 468,
'jobs': 469,
'b': 470,
'14': 471,
'18': 472,
'consumer': 473,
'manager': 474,
'lower': 475,
'department': 476,
'party': 477,
'michael': 478,
'nation': 479,
'eight': 480,
'house': 481,
'she': 482,
'saying': 483,
'northern': 484,
'energy': 485,
'japanese': 486,
'life': 487,
```

```
'europe': 488,
'center': 489,
'line': 490,
'los': 491,
'november': 492,
'southern': 493,
'running': 494,
'st': 495,
'told': 496,
'low': 497,
'champions': 498,
'little': 499,
'political': 500,
'without': 501,
'meeting': 502,
'management': 503,
'peace': 504,
'field': 505,
'due': 506,
'campaign': 507,
'hurricane': 508,
'decision': 509,
'away': 510,
'place': 511,
'taking': 512,
'launched': 513,
'central': 514,
'angeles': 515,
'russia': 516,
'good': 517,
'canadian': 518,
'winning': 519,
'oracle': 520,
'results': 521,
'florida': 522,
'race': 523,
'13': 524,
'university': 525,
'across': 526,
'school': 527,
'office': 528,
'using': 529,
'intel': 530,
'interest': 531,
'16': 532,
'star': 533,
'25': 534,
```

```
'board': 535,
'pakistan': 536,
'give': 537,
'become': 538,
'we': 539,
"world's": 540,
'net': 541,
'men': 542,
'straight': 543,
'army': 544,
'future': 545,
'morning': 546,
'secretary': 547,
'chairman': 548,
'defense': 549,
'information': 550,
'california': 551,
'50': 552,
'100': 553,
'match': 554,
'bomb': 555,
'site': 556,
'shot': 557,
'weekend': 558,
'quarterly': 559,
'costs': 560,
'launch': 561,
'lost': 562,
'charges': 563,
'communications': 564,
'linux': 565,
'case': 566,
'elections': 567,
'history': 568,
'houston': 569,
'keep': 570,
'operating': 571,
'oct': 572,
'showed': 573,
'health': 574,
'return': 575,
'better': 576,
'sun': 577,
'own': 578,
'did': 579,
'club': 580,
'same': 581,
```

```
'times': 582,
'warned': 583,
'among': 584,
'few': 585,
'posted': 586,
'signed': 587,
'increase': 588,
'death': 589,
'cost': 590,
'senior': 591,
'show': 592,
'france': 593,
'job': 594,
'mark': 595,
'television': 596,
'trading': 597,
'money': 598,
'stores': 599,
'bill': 600,
'those': 601,
'24': 602,
'peoplesoft': 603,
'earlier': 604,
'almost': 605,
'grand': 606,
'militants': 607,
'exchange': 608,
'pc': 609,
'ever': 610,
'yankees': 611,
'come': 612,
'force': 613,
'small': 614,
'david': 615,
'east': 616,
'rise': 617,
'stadium': 618,
'font': 619,
'mail': 620,
'less': 621,
'chip': 622,
'accused': 623,
'19': 624,
'nasa': 625,
'began': 626,
'nine': 627,
'ltd': 628,
```

```
'fans': 629,
'action': 630,
'iran': 631,
'going': 632,
'toronto': 633,
'likely': 634,
'nov': 635,
'texas': 636,
'reserve': 637,
'association': 638,
'baseball': 639,
'came': 640,
'thousands': 641,
'behind': 642,
'ended': 643,
'helped': 644,
'real': 645,
'barrel': 646,
'anti': 647,
'opening': 648,
'offering': 649,
'israel': 650,
'got': 651,
'full': 652,
'battle': 653,
'britain': 654,
'post': 655,
'possible': 656,
'21': 657,
'control': 658,
'amid': 659,
'died': 660,
'sell': 661,
'rate': 662,
'trial': 663,
'growing': 664,
'm': 665,
'division': 666,
'too': 667,
'august': 668,
'cash': 669,
'judge': 670,
'already': 671,
'point': 672,
'korea': 673,
'darfur': 674,
'access': 675,
```

```
'yet': 676,
'paris': 677,
'medal': 678,
'term': 679,
'food': 680,
'tax': 681,
'performance': 682,
'members': 683,
'hopes': 684,
'hours': 685,
'kerry': 686,
'bankruptcy': 687,
'countries': 688,
'far': 689,
'white': 690,
'once': 691,
'women': 692,
'administration': 693,
'sept': 694,
'revenue': 695,
'career': 696,
'fall': 697,
'151': 698,
'injured': 699,
'rates': 700,
'networks': 701,
'trying': 702,
'western': 703,
'production': 704,
'gave': 705,
'store': 706,
'study': 707,
'efforts': 708,
'airlines': 709,
'chinese': 710,
'within': 711,
'see': 712,
'authorities': 713,
'soon': 714,
'violence': 715,
'yards': 716,
'williams': 717,
'product': 718,
'computers': 719,
'german': 720,
'corporate': 721,
'hard': 722,
```

```
'look': 723,
'germany': 724,
'soldiers': 725,
'local': 726,
'george': 727,
'speed': 728,
'human': 729,
'radio': 730,
'server': 731,
'hostage': 732,
'takeover': 733,
'step': 734,
'canada': 735,
'change': 736,
'main': 737,
'starting': 738,
'popular': 739,
'opposition': 740,
'personal': 741,
'outside': 742,
'competition': 743,
'paul': 744,
'aimed': 745,
'each': 746,
'quarterback': 747,
'markets': 748,
'fund': 749,
'reached': 750,
'investment': 751,
'african': 752,
'fire': 753,
'size': 754,
'law': 755,
'until': 756,
'arafat': 757,
'short': 758,
'concerns': 759,
'strike': 760,
'engine': 761,
'known': 762,
'continued': 763,
'phones': 764,
'failed': 765,
'jones': 766,
'vote': 767,
'enough': 768,
'devices': 769,
```

```
'giants': 770,
'yasser': 771,
'retailer': 772,
'gas': 773,
'afghanistan': 774,
'then': 775,
'analysts': 776,
'taken': 777,
'begin': 778,
'others': 779,
'development': 780,
'tour': 781,
'profits': 782,
'looking': 783,
'dead': 784,
'manchester': 785,
'filed': 786,
'weapons': 787,
'tech': 788,
'fight': 789,
'airline': 790,
'insurance': 791,
'selling': 792,
'stop': 793,
'reach': 794,
'council': 795,
'designed': 796,
'unit': 797,
'miami': 798,
'raised': 799,
'seattle': 800,
'desktop': 801,
'large': 802,
'teams': 803,
'sony': 804,
'cell': 805,
'americans': 806,
'bay': 807,
'beijing': 808,
'side': 809,
're': 810,
'electronics': 811,
'executives': 812,
'meet': 813,
'indian': 814,
'great': 815,
'olympics': 816,
```

```
'find': 817,
'500': 818,
"it's": 819,
'went': 820,
'legal': 821,
'must': 822,
'effort': 823,
'hold': 824,
'rebel': 825,
'operations': 826,
'fuel': 827,
'further': 828,
'finally': 829,
'cp': 830,
'green': 831,
'fighting': 832,
'tony': 833,
'powerful': 834,
'coming': 835,
'consumers': 836,
'researchers': 837,
'working': 838,
'level': 839,
'spending': 840,
'file': 841,
'coast': 842,
'call': 843,
'moscow': 844,
'need': 845,
'rights': 846,
'area': 847,
'toward': 848,
'fired': 849,
'madrid': 850,
'claims': 851,
'atlanta': 852,
'family': 853,
'annual': 854,
'forecast': 855,
'getting': 856,
'securities': 857,
'continue': 858,
'businesses': 859,
'23': 860,
'allow': 861,
'rally': 862,
'private': 863,
```

```
'never': 864,
'euro': 865,
'arrested': 866,
'station': 867,
'goal': 868,
'tv': 869,
'list': 870,
'bowl': 871,
'blue': 872,
"'s": 873,
'heart': 874,
'yahoo': 875,
'every': 876,
'labor': 877,
'provide': 878,
'cbs': 879,
'pressure': 880,
'rebels': 881,
'democratic': 882,
'africa': 883,
'27': 884,
'visit': 885,
'planned': 886,
'children': 887,
'calif': 888,
'italian': 889,
'holiday': 890,
'investigation': 891,
'basketball': 892,
'drop': 893,
'drive': 894,
'name': 895,
'sent': 896,
'spain': 897,
'available': 898,
'boost': 899,
'hundreds': 900,
'survey': 901,
'28': 902,
'current': 903,
'charged': 904,
'regulators': 905,
'attempt': 906,
'named': 907,
'dollars': 908,
'officer': 909,
'expectations': 910,
```

```
'policy': 911,
'muslim': 912,
'rising': 913,
'22': 914,
'college': 915,
'problems': 916,
'sudan': 917,
'asked': 918,
'suspected': 919,
'approved': 920,
'ipod': 921,
'sold': 922,
'nasdaq': 923,
'detroit': 924,
'december': 925,
'gains': 926,
'decided': 927,
'aid': 928,
'supply': 929,
'ceo': 930,
'park': 931,
'employees': 932,
'philadelphia': 933,
'retail': 934,
'nfl': 935,
'flight': 936,
'unveiled': 937,
'worth': 938,
'islamic': 939,
'confirmed': 940,
'tournament': 941,
'arsenal': 942,
'turned': 943,
'seen': 944,
'rivals': 945,
'sale': 946,
'eu': 947,
'equipment': 948,
'louis': 949,
'injury': 950,
'strip': 951,
'order': 952,
'claimed': 953,
'sign': 954,
'holding': 955,
'ruling': 956,
'later': 957,
```

```
'technologies': 958,
'town': 959,
'winter': 960,
'increased': 961,
'try': 962,
'terrorism': 963,
'giving': 964,
'color': 965,
'staff': 966,
'moved': 967,
'spokesman': 968,
'stake': 969,
'want': 970,
'guard': 971,
'fresh': 972,
'though': 973,
'40': 974,
'opened': 975,
'mike': 976,
'fifth': 977,
'generation': 978,
'above': 979,
'cent': 980,
'average': 981,
'w': 982,
'received': 983,
'threat': 984,
'inning': 985,
'chance': 986,
'double': 987,
'hospital': 988,
'touchdown': 989,
'added': 990,
'ranked': 991,
'steve': 992,
'2006': 993,
'crisis': 994,
'process': 995,
'applications': 996,
'captain': 997,
'2003': 998,
'suicide': 999,
'black': 1000,
…}
```

```
[21]: sequences = tokenizer.texts_to_sequences(news)
```

```
[22]: index = 1
      print(news[index])
      print(sequences[index])
```

Reuters - Private investment firm Carlyle Group,\which has a reputation for
making well-timed and occasionally\controversial plays in the defense industry,
has quietly placed\its bets on another part of the market.
[24, 863, 751, 371, 93, 84, 20, 2, 3916, 8, 453, 431, 6, 1308, 2799, 5, 1, 549,
237, 20, 3528, 2002, 14, 8267, 7, 216, 314, 4, 1, 131]

```
[23]: tokenizer.word_index['investment']
```

```
[23]: 751
```

```
[24]: x_train = pad_sequences(sequences, maxlen=max_news_len)
```

```
[25]: x_train[:5]
```

```
[25]: array([[    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
                 0,    0,    0,    0,    0,    0,    0,   24,  758, 7851,  433,
              5786, 2861,    4, 5916,   33, 3642,  831,  432],
            [   24,  863,  751,  371,   93,   84,   20,    2, 3916,    8,  453,
               431,    6, 1308, 2799,    5,    1,  549,  237,   20, 3528, 2002,
                14, 8267,    7,  216,  314,    4,    1,  131],
            [   24, 2199,  463,  105, 1568, 1484,   64,    1,  397,    6,    1,
              1026,    8,  317,   33,  178,    3, 6377,   35,    1,  311,  131,
                99,   79,  189,    1, 6120,    4,    1, 1068],
            [    0,   24,  713,   29, 5142,   77, 3549, 7993,   23,    1,  737,
              3199,    5,  493,  106,   28, 1402,  573,    2,  825, 2601,   90,
               760, 2559,   22,   77,  292,   16,    7,   97],
            [    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,  165,
                55,   77,  105, 8851, 1776,    6, 8268, 3454,    2,   21,  343,
              3036,   70,  266,  151,    1,   34,  340,  567]], dtype=int32)
```

0.6

```
[26]: model_cnn = Sequential()
      model_cnn.add(Embedding(num_words, 32, input_length=max_news_len))
      model_cnn.add(Conv1D(250, 5, padding='valid', activation='relu'))
      model_cnn.add(GlobalMaxPooling1D())
      model_cnn.add(Dense(128, activation='relu'))
      model_cnn.add(Dense(4, activation='softmax'))
```

```
/usr/local/lib/python3.10/dist-packages/keras/src/layers/core/embedding.py:90:
UserWarning: Argument `input_length` is deprecated. Just remove it.
  warnings.warn(
```

[27]:
```python
model_cnn.compile(optimizer='adam',
                  loss='categorical_crossentropy',
                  metrics=['accuracy'])
```

[28]:
```python
model_cnn.summary()
```

Model: "sequential"

Layer (type)	Output Shape	␣ ↪Param #
embedding (Embedding) ↪(unbuilt)	?	0␣
conv1d (Conv1D) ↪(unbuilt)	?	0␣
global_max_pooling1d ↪(unbuilt) (GlobalMaxPooling1D) ↪	?	0␣ ␣
dense (Dense) ↪(unbuilt)	?	0␣
dense_1 (Dense) ↪(unbuilt)	?	0␣

 Total params: 0 (0.00 B)

 Trainable params: 0 (0.00 B)

 Non-trainable params: 0 (0.00 B)

     callback   ,   .
   best_model.h5
```
```

```python
[29]: model_cnn_save_path = 'best_model_cnn.h5'
      checkpoint_callback_cnn = ModelCheckpoint(model_cnn_save_path,
                                                monitor='val_accuracy',
                                                save_best_only=True,
                                                verbose=1)
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
<ipython-input-29-f3b35d1108ac> in <cell line: 2>()
      1 model_cnn_save_path = 'best_model_cnn.h5'
----> 2 checkpoint_callback_cnn = ModelCheckpoint(model_cnn_save_path,
      3                                            monitor='val_accuracy',
      4                                            save_best_only=True,
      5                                            verbose=1)

/usr/local/lib/python3.10/dist-packages/keras/src/callbacks/model_checkpoint.py
  ↪in __init__(self, filepath, monitor, verbose, save_best_only,␣
  ↪save_weights_only, mode, save_freq, initial_value_threshold)
    189         else:
    190             if not self.filepath.endswith(".keras"):
--> 191                 raise ValueError(
    192                     "The filepath provided must end in `.keras` "
    193                     "(Keras model format). Received: "

ValueError: The filepath provided must end in `.keras` (Keras model format).␣
  ↪Received: filepath=best_model_cnn.h5
```

```python
[ ]: history_cnn = model_cnn.fit(x_train,
                                  y_train,
                                  epochs=5,
                                  batch_size=128,
                                  validation_split=0.1,
                                  callbacks=[checkpoint_callback_cnn])
```

```python
[ ]: plt.plot(history_cnn.history['accuracy'],
             label='                      ')
     plt.plot(history_cnn.history['val_accuracy'],
             label='                      ')
     plt.xlabel('          ')
     plt.ylabel('           ')
     plt.legend()
     plt.show()
```

0.7 C LSTM

```python
model_lstm = Sequential()
model_lstm.add(Embedding(num_words, 32, input_length=max_news_len))
model_lstm.add(LSTM(16))
model_lstm.add(Dense(4, activation='softmax'))
```

```python
model_lstm.compile(optimizer='adam',
                   loss='categorical_crossentropy',
                   metrics=['accuracy'])
```

```python
model_lstm.summary()
```

callback   ,   .
best_model.h5

```python
model_lstm_save_path = 'best_model_lstm.h5'
checkpoint_callback_lstm = ModelCheckpoint(model_lstm_save_path,
                                           monitor='val_accuracy',
                                           save_best_only=True,
                                           verbose=1)
```

```python
history_lstm = model_lstm.fit(x_train,
                              y_train,
                              epochs=5,
                              batch_size=128,
                              validation_split=0.1,
                              callbacks=[checkpoint_callback_lstm])
```

```python
plt.plot(history_lstm.history['accuracy'],
         label='                              ')
plt.plot(history_lstm.history['val_accuracy'],
         label='                              ')
plt.xlabel('          ')
plt.ylabel('             ')
plt.legend()
plt.show()
```

0.8 C GRU

```python
model_gru = Sequential()
model_gru.add(Embedding(num_words, 32, input_length=max_news_len))
model_gru.add(GRU(16))
model_gru.add(Dense(4, activation='softmax'))
```

```python
model_gru.compile(optimizer='adam',
                  loss='categorical_crossentropy',
                  metrics=['accuracy'])
```

```
model_gru.summary()
```

callback   ,   .
best_model_gru.h5

```
model_gru_save_path = 'best_model_gru.h5'
checkpoint_callback_gru = ModelCheckpoint(model_gru_save_path,
                                          monitor='val_accuracy',
                                          save_best_only=True,
                                          verbose=1)
```

```
history_gru = model_gru.fit(x_train,
                            y_train,
                            epochs=5,
                            batch_size=128,
                            validation_split=0.1,
                            callbacks=[checkpoint_callback_gru])
```

```
plt.plot(history_gru.history['accuracy'],
         label='                        ')
plt.plot(history_gru.history['val_accuracy'],
         label='                          ')
plt.xlabel('          ')
plt.ylabel('           ')
plt.legend()
plt.show()
```

0.9

```
test = pd.read_csv('test.csv',
                   header=None,
                   names=['class', 'title', 'text'])
```

```
test
```

```
test_sequences = tokenizer.texts_to_sequences(test['text'])
```

```
x_test = pad_sequences(test_sequences, maxlen=max_news_len)
```

```
x_test[:5]
```

```
y_test = utils.to_categorical(test['class'] - 1, nb_classes)
```

```
y_test
```

0.10

0.10.1

```
[ ]: model_cnn.load_weights(model_cnn_save_path)
```

```
[ ]: model_cnn.evaluate(x_test, y_test, verbose=1)
```

0.10.2 LSTM

```
[ ]: model_lstm.load_weights(model_lstm_save_path)
```

```
[ ]: model_lstm.evaluate(x_test, y_test, verbose=1)
```

0.10.3 GRU

```
[ ]: model_gru.load_weights(model_gru_save_path)
```

```
[ ]: model_gru.evaluate(x_test, y_test, verbose=1)
```

```
[ ]: ###
```

```
[ ]: model.save('news.h5')
```

```
[ ]:
```