LSTM

August 4, 2025

0.1 Project Description: Next Word Prediction Using LSTM

Project Overview: This project aims to develop a deep learning model for predicting the next word in a given sequence of words. The model is built using Long Short-Term Memory (LSTM) networks, which are well-suited for sequence prediction tasks. The project includes the following steps:

1- Data Collection: We use the text of Shakespeare's "Hamlet" as our dataset. This rich, complex text provides a good challenge for our model.

2- Data Preprocessing: The text data is tokenized, converted into sequences, and padded to ensure uniform input lengths. The sequences are then split into training and testing sets.

3- Model Building: An LSTM model is constructed with an embedding layer, two LSTM layers, and a dense output layer with a softmax activation function to predict the probability of the next word.

4- Model Training: The model is trained using the prepared sequences, with early stopping implemented to prevent overfitting. Early stopping monitors the validation loss and stops training when the loss stops improving.

5- Model Evaluation: The model is evaluated using a set of example sentences to test its ability to predict the next word accurately.

6- Deployment: A Streamlit web application is developed to allow users to input a sequence of words and get the predicted next word in real-time.

[46]:
```
## Data Collection
import nltk
nltk.download('gutenberg')
from nltk.corpus import gutenberg
import pandas as pd

## load the dataset
data = gutenberg.raw('shakespeare-hamlet.txt')
## save to a file
with open('hamlet.txt','w') as file:
  file.write(data)
```

```
[nltk_data] Downloading package gutenberg to /root/nltk_data…
[nltk_data]   Package gutenberg is already up-to-date!
```

[47]:
```
## Data Preprocessing
import numpy as np
from tensorflow.keras.preprocessing.text import Tokenizer
from tensorflow.keras.preprocessing.sequence import pad_sequences
from sklearn.model_selection import train_test_split

##laod the dataset

with open('hamlet.txt','r') as file:
  text = file.read().lower()

## Tokenize the text-creating indexes for words

tokenizer = Tokenizer()
tokenizer.fit_on_texts([text])
total_words = len(tokenizer.word_index) + 1
total_words
```

[47]: 4818

[48]:
```
tokenizer.word_index
```

[48]:
```
{'the': 1,
 'and': 2,
 'to': 3,
 'of': 4,
 'i': 5,
 'you': 6,
 'a': 7,
 'my': 8,
 'it': 9,
 'in': 10,
 'that': 11,
 'ham': 12,
 'is': 13,
 'not': 14,
 'his': 15,
 'this': 16,
 'with': 17,
 'your': 18,
 'but': 19,
 'for': 20,
 'me': 21,
 'lord': 22,
 'as': 23,
 'what': 24,
 'he': 25,
```

```
'be': 26,
'so': 27,
'him': 28,
'haue': 29,
'king': 30,
'will': 31,
'no': 32,
'our': 33,
'we': 34,
'on': 35,
'are': 36,
'if': 37,
'all': 38,
'then': 39,
'shall': 40,
'by': 41,
'thou': 42,
'come': 43,
'or': 44,
'hamlet': 45,
'good': 46,
'do': 47,
'hor': 48,
'her': 49,
'let': 50,
'now': 51,
'thy': 52,
'how': 53,
'more': 54,
'they': 55,
'from': 56,
'enter': 57,
'at': 58,
'was': 59,
'oh': 60,
'like': 61,
'most': 62,
'there': 63,
'well': 64,
'know': 65,
'selfe': 66,
'would': 67,
'them': 68,
'loue': 69,
'may': 70,
"'tis": 71,
'vs': 72,
```

```
'sir': 73,
'qu': 74,
'which': 75,
'did': 76,
'why': 77,
'laer': 78,
'giue': 79,
'thee': 80,
'ile': 81,
'must': 82,
'hath': 83,
'ophe': 84,
'speake': 85,
'out': 86,
'make': 87,
'should': 88,
'where': 89,
'too': 90,
'an': 91,
'am': 92,
'such': 93,
'say': 94,
'when': 95,
'vpon': 96,
'father': 97,
'doe': 98,
'very': 99,
'pol': 100,
'go': 101,
'their': 102,
'one': 103,
'man': 104,
'see': 105,
'some': 106,
'heere': 107,
'had': 108,
'heauen': 109,
'time': 110,
'mine': 111,
'these': 112,
'she': 113,
'much': 114,
'tell': 115,
'rosin': 116,
'thinke': 117,
'play': 118,
'thus': 119,
```

```
'horatio': 120,
'who': 121,
'mother': 122,
'queene': 123,
'night': 124,
'o': 125,
'polon': 126,
'yet': 127,
'vp': 128,
'death': 129,
'laertes': 130,
'againe': 131,
'can': 132,
'both': 133,
"th'": 134,
'soule': 135,
'take': 136,
'life': 137,
'nor': 138,
'heare': 139,
'mar': 140,
'looke': 141,
'owne': 142,
'could': 143,
'heart': 144,
'dead': 145,
'might': 146,
'made': 147,
'clo': 148,
'hast': 149,
'downe': 150,
'pray': 151,
'ophelia': 152,
'nothing': 153,
'away': 154,
'whose': 155,
'doth': 156,
'other': 157,
'cannot': 158,
'leaue': 159,
'indeed': 160,
'into': 161,
'nay': 162,
'god': 163,
'head': 164,
'were': 165,
'matter': 166,
```

```
'thing': 167,
'hold': 168,
'day': 169,
'world': 170,
'nature': 171,
'neuer': 172,
'comes': 173,
'done': 174,
'exeunt': 175,
'call': 176,
'two': 177,
'true': 178,
'though': 179,
'sweet': 180,
'put': 181,
'set': 182,
'ghost': 183,
'euen': 184,
'earth': 185,
'feare': 186,
'madnesse': 187,
'mad': 188,
'seene': 189,
'eyes': 190,
'against': 191,
'faire': 192,
'denmarke': 193,
'those': 194,
"o're": 195,
'polonius': 196,
'deere': 197,
'fathers': 198,
'sonne': 199,
'poore': 200,
'himselfe': 201,
'follow': 202,
'guild': 203,
'england': 204,
'friends': 205,
'once': 206,
'hand': 207,
'shew': 208,
'about': 209,
"i'th'": 210,
'off': 211,
'within': 212,
'till': 213,
```

```
'great': 214,
'meanes': 215,
'words': 216,
'players': 217,
'exit': 218,
'part': 219,
'still': 220,
'does': 221,
'hee': 222,
'osr': 223,
'long': 224,
'before': 225,
'beleeue': 226,
'any': 227,
'old': 228,
'thoughts': 229,
'first': 230,
'eare': 231,
'keepe': 232,
'goe': 233,
'end': 234,
'guildensterne': 235,
'welcome': 236,
'while': 237,
'art': 238,
'noble': 239,
'body': 240,
'bee': 241,
'daughter': 242,
'speech': 243,
'makes': 244,
"there's": 245,
'sword': 246,
'stand': 247,
'liue': 248,
"that's": 249,
'farewell': 250,
'kin': 251,
'ere': 252,
'marry': 253,
'betweene': 254,
'many': 255,
'since': 256,
'watch': 257,
"ha's": 258,
'therefore': 259,
'question': 260,
```

```
'thought': 261,
'heard': 262,
'spirit': 263,
'eye': 264,
'better': 265,
'thine': 266,
'tongue': 267,
'drinke': 268,
'youth': 269,
'sent': 270,
'graue': 271,
'rest': 272,
'bed': 273,
'last': 274,
'same': 275,
'marke': 276,
'gone': 277,
'without': 278,
'state': 279,
"is't": 280,
'goes': 281,
'fortinbras': 282,
'vse': 283,
'grace': 284,
'euer': 285,
'finde': 286,
'gertrude': 287,
'beare': 288,
'little': 289,
'breath': 290,
"wee'l": 291,
'saw': 292,
'beene': 293,
'none': 294,
'vertue': 295,
'else': 296,
'said': 297,
'after': 298,
'reynol': 299,
'cause': 300,
'forme': 301,
'something': 302,
'ayre': 303,
'farre': 304,
'selues': 305,
'purpose': 306,
'further': 307,
```

```
'reason': 308,
'friend': 309,
'madam': 310,
'remember': 311,
'faith': 312,
'gentlemen': 313,
'word': 314,
'foule': 315,
'winde': 316,
'meane': 317,
'bring': 318,
'fit': 319,
'blood': 320,
'helpe': 321,
'honest': 322,
'stay': 323,
"in't": 324,
'being': 325,
'fire': 326,
'things': 327,
"what's": 328,
'newes': 329,
'best': 330,
'kinde': 331,
'excellent': 332,
'each': 333,
'sleepe': 334,
'way': 335,
'please': 336,
'free': 337,
'reuenge': 338,
'villaine': 339,
'right': 340,
'ha': 341,
'passion': 342,
'rosincrance': 343,
'dost': 344,
'verie': 345,
'barn': 346,
'marcellus': 347,
'men': 348,
'peace': 349,
'together': 350,
'full': 351,
'voyce': 352,
'oft': 353,
'greefe': 354,
```

```
"'twere": 355,
'late': 356,
'businesse': 357,
'doubt': 358,
'alone': 359,
'minde': 360,
'heauens': 361,
'face': 362,
'hell': 363,
'ye': 364,
'second': 365,
'iudgement': 366,
'giuen': 367,
'command': 368,
'action': 369,
"let's": 370,
'murther': 371,
'guil': 372,
'lady': 373,
'fortune': 374,
'mee': 375,
'pyrrhus': 376,
'answer': 377,
'get': 378,
'thankes': 379,
'goodnight': 380,
'eares': 381,
'breake': 382,
'hora': 383,
'strange': 384,
'young': 385,
'walke': 386,
'brothers': 387,
'seeme': 388,
'name': 389,
'fellow': 390,
'act': 391,
'hands': 392,
'armes': 393,
'deare': 394,
'neere': 395,
'phrase': 396,
'draw': 397,
'gho': 398,
'alas': 399,
'ought': 400,
'offence': 401,
```

```
'sweare': 402,
'worke': 403,
'gentleman': 404,
'fine': 405,
'three': 406,
'barnardo': 407,
'fran': 408,
'ground': 409,
'sight': 410,
'sit': 411,
'maiesty': 412,
'pale': 413,
"on't": 414,
'fell': 415,
'lost': 416,
'soft': 417,
'power': 418,
'yong': 419,
'duty': 420,
'whole': 421,
'woe': 422,
'ioy': 423,
'wife': 424,
'came': 425,
'queen': 426,
'seeke': 427,
'common': 428,
'seemes': 429,
'blacke': 430,
'kings': 431,
'teares': 432,
'top': 433,
'fashion': 434,
'deed': 435,
'euery': 436,
'light': 437,
'custome': 438,
'borne': 439,
'wilt': 440,
'hither': 441,
'lay': 442,
'another': 443,
'ouer': 444,
'age': 445,
'thousand': 446,
'fall': 447,
'lye': 448,
```

```
'conscience': 449,
'husband': 450,
'bar': 451,
'lookes': 452,
'charge': 453,
'knowne': 454,
'law': 455,
'bin': 456,
'sound': 457,
'sister': 458,
'memory': 459,
'brother': 460,
'beseech': 461,
'lesse': 462,
'dust': 463,
'through': 464,
'shewes': 465,
'desire': 466,
'obey': 467,
'woman': 468,
'almost': 469,
'grow': 470,
'here': 471,
'shame': 472,
'giues': 473,
"too't": 474,
'takes': 475,
'table': 476,
'sure': 477,
'musicke': 478,
'letters': 479,
'hamlets': 480,
'hope': 481,
'receiue': 482,
'maiestie': 483,
'thanke': 484,
'gaue': 485,
'bad': 486,
'wee': 487,
'ore': 488,
'noise': 489,
'times': 490,
'cold': 491,
'bid': 492,
'dane': 493,
'place': 494,
'peece': 495,
```

```
'buried': 496,
'cast': 497,
'hot': 498,
'list': 499,
'wrong': 500,
'sea': 501,
'truth': 502,
'sayes': 503,
'season': 504,
'gracious': 505,
'dumbe': 506,
'loues': 507,
'sorrow': 508,
'marriage': 509,
'writ': 510,
'mouth': 511,
'pardon': 512,
'note': 513,
'backe': 514,
'lordship': 515,
'mothers': 516,
'beard': 517,
'fare': 518,
'seruice': 519,
'withall': 520,
'maid': 521,
'enough': 522,
'effect': 523,
'double': 524,
'neither': 525,
'false': 526,
'vnderstand': 527,
'circumstance': 528,
'foole': 529,
'vowes': 530,
'keepes': 531,
'shape': 532,
'dayes': 533,
'fat': 534,
'crowne': 535,
'wits': 536,
'damned': 537,
'ho': 538,
'needs': 539,
'touch': 540,
'ranke': 541,
'generall': 542,
```

```
'moue': 543,
'home': 544,
'ill': 545,
'round': 546,
'fortunes': 547,
'laugh': 548,
'yours': 549,
"he's": 550,
'honor': 551,
'begin': 552,
'anon': 553,
'proofe': 554,
'gods': 555,
'quicke': 556,
'dangerous': 557,
'christian': 558,
'danish': 559,
'poyson': 560,
'begge': 561,
'wager': 562,
"drown'd": 563,
'water': 564,
'scull': 565,
'houre': 566,
'twelue': 567,
'quiet': 568,
'course': 569,
'sometimes': 570,
'march': 571,
'look': 572,
'norwey': 573,
'particular': 574,
'land': 575,
"do's": 576,
'vnto': 577,
'speak': 578,
'spirits': 579,
'cocke': 580,
'guilty': 581,
"'gainst": 582,
'wholsome': 583,
'lords': 584,
'kingdome': 585,
'freely': 586,
'dreame': 587,
'told': 588,
'loose': 589,
```

```
'dread': 590,
'returne': 591,
'france': 592,
'confesse': 593,
'dye': 594,
'visage': 595,
'truly': 596,
'bound': 597,
'prythee': 598,
'health': 599,
'flesh': 600,
'fie': 601,
'beast': 602,
'discourse': 603,
'longer': 604,
'wicked': 605,
'disposition': 606,
'report': 607,
'teach': 608,
'forth': 609,
'thinkes': 610,
'tis': 611,
'yes': 612,
'countenance': 613,
'perchance': 614,
'warrant': 615,
'silence': 616,
'perhaps': 617,
'wisedome': 618,
'blessing': 619,
'dull': 620,
'mans': 621,
'audience': 622,
"you'l": 623,
'making': 624,
"damn'd": 625,
'soules': 626,
'cries': 627,
'desperate': 628,
'shalt': 629,
'prison': 630,
'went': 631,
'naturall': 632,
'holds': 633,
'sodaine': 634,
'adue': 635,
'braine': 636,
```

```
'knaue': 637,
'point': 638,
'vnder': 639,
'mercy': 640,
'lacke': 641,
"heere's": 642,
"in's": 643,
'tooke': 644,
'arme': 645,
'brought': 646,
'dutie': 647,
'found': 648,
'whereon': 649,
'commission': 650,
'passe': 651,
'vilde': 652,
'short': 653,
'try': 654,
'behinde': 655,
'presently': 656,
'slaue': 657,
'saue': 658,
'ambition': 659,
'sing': 660,
'themselues': 661,
'braines': 662,
"'twas": 663,
'onely': 664,
'french': 665,
'treason': 666,
'morrow': 667,
'conceit': 668,
'drowne': 669,
'fellowes': 670,
'hoa': 671,
'patience': 672,
'halfe': 673,
'yeare': 674,
'forgot': 675,
'hence': 676,
"doo't": 677,
"e'ene": 678,
'slaine': 679,
'sense': 680,
'buriall': 681,
'alexander': 682,
'osricke': 683,
```

```
'carriages': 684,
'foyles': 685,
'hit': 686,
'tragedie': 687,
'sicke': 688,
'meet': 689,
'twice': 690,
"'twill": 691,
'nights': 692,
'starre': 693,
"t'": 694,
'figure': 695,
'wonder': 696,
'warlike': 697,
'cannon': 698,
'toward': 699,
'image': 700,
'norway': 701,
"seal'd": 702,
'lands': 703,
'stood': 704,
"return'd": 705,
'strong': 706,
'motiue': 707,
'ease': 708,
'treasure': 709,
'stop': 710,
'violence': 711,
'trumpet': 712,
'whether': 713,
'heerein': 714,
'wherein': 715,
'impart': 716,
'morning': 717,
'attendant': 718,
'brow': 719,
'discretion': 720,
'followes': 721,
'frame': 722,
'cornelius': 723,
'giuing': 724,
'commend': 725,
"would'st": 726,
'natiue': 727,
'fauour': 728,
'hang': 729,
'colour': 730,
```

```
'show': 731,
'fault': 732,
'throw': 733,
'lose': 734,
'gentle': 735,
'sits': 736,
'growes': 737,
'visit': 738,
'growne': 739,
'vnkle': 740,
'left': 741,
'speed': 742,
'incestuous': 743,
'glad': 744,
'forget': 745,
'thrift': 746,
"arm'd": 747,
'thrice': 748,
'length': 749,
'kept': 750,
'knew': 751,
'answere': 752,
'honour': 753,
'wide': 754,
'force': 755,
'affection': 756,
'shot': 757,
'danger': 758,
'moone': 759,
'safety': 760,
'lies': 761,
'buy': 762,
'aboue': 763,
'humbly': 764,
'tend': 765,
'ist': 766,
'bloud': 767,
'heate': 768,
'bones': 769,
'base': 770,
'horrible': 771,
'imagination': 772,
'direct': 773,
'lend': 774,
'hearing': 775,
'certaine': 776,
'fast': 777,
```

```
'house': 778,
'tale': 779,
'start': 780,
'vnnaturall': 781,
"it's": 782,
'sleeping': 783,
'wit': 784,
'gifts': 785,
'seeming': 786,
'wil': 787,
'court': 788,
'cursed': 789,
'instant': 790,
'bosome': 791,
'distracted': 792,
'yea': 793,
'past': 794,
'booke': 795,
'think': 796,
'secret': 797,
'already': 798,
'stage': 799,
'seeing': 800,
'fingers': 801,
'reynoldo': 802,
'drift': 803,
'liberty': 804,
'closes': 805,
'consequence': 806,
'chamber': 807,
"turn'd": 808,
'extasie': 809,
'violent': 810,
'meant': 811,
'hide': 812,
'whom': 813,
'rather': 814,
'white': 815,
'awhile': 816,
'idle': 817,
'thence': 818,
'bene': 819,
'foure': 820,
'honestie': 821,
'either': 822,
'count': 823,
'dreames': 824,
```

```
'bodies': 825,
'comming': 826,
'shal': 827,
'player': 828,
'flourish': 829,
'asse': 830,
'heauy': 831,
'masters': 832,
'1': 833,
'others': 834,
'modestie': 835,
'cunning': 836,
"lou'd": 837,
'horse': 838,
'priam': 839,
'new': 840,
'sleepes': 841,
'hecuba': 842,
'mortall': 843,
'weepe': 844,
'pate': 845,
'diuell': 846,
'blame': 847,
'flye': 848,
'turne': 849,
'beautie': 850,
'acte': 851,
'nunnery': 852,
'snow': 853,
'rose': 854,
'quite': 855,
'ladies': 856,
'wretched': 857,
'send': 858,
'weare': 859,
"kill'd": 860,
'kill': 861,
'maker': 862,
'prologue': 863,
'shortly': 864,
'begun': 865,
'lights': 866,
'heeles': 867,
'doore': 868,
'meete': 869,
'stronger': 870,
'whereto': 871,
```

```
'next': 872,
'messenger': 873,
'choose': 874,
'bore': 875,
'spade': 876,
'gallowes': 877,
'sings': 878,
'rites': 879,
"woo't": 880,
'cup': 881,
'vnfold': 882,
'bitter': 883,
'guard': 884,
'mouse': 885,
"appear'd": 886,
'saies': 887,
'touching': 888,
'along': 889,
'appeare': 890,
'beating': 891,
'offended': 892,
'ambitious': 893,
'iust': 894,
'knowes': 895,
'subiect': 896,
'sore': 897,
'least': 898,
'valiant': 899,
'recouer': 900,
'termes': 901,
'maine': 902,
'loe': 903,
'offer': 904,
'present': 905,
'dew': 906,
'high': 907,
'hill': 908,
'aduice': 909,
'consent': 910,
'scena': 911,
'greene': 912,
'hearts': 913,
'wisest': 914,
'remembrance': 915,
'funerall': 916,
'delight': 917,
'affaire': 918,
```

```
'thinking': 919,
'voltemand': 920,
'vncle': 921,
'heartily': 922,
'bend': 923,
'bow': 924,
'cosin': 925,
'clouds': 926,
'sun': 927,
'liues': 928,
'passing': 929,
'formes': 930,
'vnderstanding': 931,
'sence': 932,
'coarse': 933,
'dyed': 934,
'beares': 935,
'wittenberg': 936,
'courtier': 937,
'louing': 938,
'smiling': 939,
'manet': 940,
'fixt': 941,
'flat': 942,
'vses': 943,
'married': 944,
'hercules': 945,
'salt': 946,
'change': 947,
'deepe': 948,
'mock': 949,
'hard': 950,
'tables': 951,
'goodly': 952,
'admiration': 953,
'deliuer': 954,
'cap': 955,
'appeares': 956,
'whilst': 957,
'dreadfull': 958,
'third': 959,
'sirs': 960,
'foote': 961,
'staid': 962,
'assume': 963,
'person': 964,
'mens': 965,
```

```
 'choyce': 966,
 'beauty': 967,
 'spring': 968,
 'aboord': 969,
 'saile': 970,
 'few': 971,
 'steele': 972,
 'entertainment': 973,
 'censure': 974,
 'rich': 975,
 'generous': 976,
 'edge': 977,
 'tenders': 978,
 'pay': 979,
 'tender': 980,
 'honourable': 981,
 'presence': 982,
 'meere': 983,
 'wayes': 984,
 'hower': 985,
 'strooke': 986,
 'wont': 987,
 "honour'd": 988,
 'angels': 989,
 'royall': 990,
 'burst': 991,
 'ignorance': 992,
 'shake': 993,
 'wherefore': 994,
 'tempt': 995,
 'lets': 996,
 'lead': 997,
 'crimes': 998,
 'starres': 999,
 'stirre': 1000,
 …}
```

[49]:
```
## create inoput sequences
input_sequences = []
for line in text.split('\n'):
  token_list = tokenizer.texts_to_sequences([line])[0]
  for i in range (1, len(token_list)):
    n_gram_sequence = token_list[:i+1]
    input_sequences.append(n_gram_sequence)
```

[50]:
```
input_sequences
```

```
[50]: [[1, 687],
       [1, 687, 4],
       [1, 687, 4, 45],
       [1, 687, 4, 45, 41],
       [1, 687, 4, 45, 41, 1886],
       [1, 687, 4, 45, 41, 1886, 1887],
       [1, 687, 4, 45, 41, 1886, 1887, 1888],
       [1180, 1889],
       [1180, 1889, 1890],
       [1180, 1889, 1890, 1891],
       [57, 407],
       [57, 407, 2],
       [57, 407, 2, 1181],
       [57, 407, 2, 1181, 177],
       [57, 407, 2, 1181, 177, 1892],
       [407, 1182],
       [407, 1182, 63],
       [408, 162],
       [408, 162, 377],
       [408, 162, 377, 21],
       [408, 162, 377, 21, 247],
       [408, 162, 377, 21, 247, 882],
       [18, 66],
       [451, 224],
       [451, 224, 248],
       [451, 224, 248, 1],
       [451, 224, 248, 1, 30],
       [408, 407],
       [451, 25],
       [408, 6],
       [408, 6, 43],
       [408, 6, 43, 62],
       [408, 6, 43, 62, 1893],
       [408, 6, 43, 62, 1893, 96],
       [408, 6, 43, 62, 1893, 96, 18],
       [408, 6, 43, 62, 1893, 96, 18, 566],
       [451, 71],
       [451, 71, 51],
       [451, 71, 51, 1894],
       [451, 71, 51, 1894, 567],
       [451, 71, 51, 1894, 567, 378],
       [451, 71, 51, 1894, 567, 378, 80],
       [451, 71, 51, 1894, 567, 378, 80, 3],
       [451, 71, 51, 1894, 567, 378, 80, 3, 273],
       [451, 71, 51, 1894, 567, 378, 80, 3, 273, 1181],
       [408, 20],
       [408, 20, 16],
```

```
[408, 20, 16, 1895],
[408, 20, 16, 1895, 114],
[408, 20, 16, 1895, 114, 379],
[408, 20, 16, 1895, 114, 379, 71],
[408, 20, 16, 1895, 114, 379, 71, 883],
[408, 20, 16, 1895, 114, 379, 71, 883, 491],
[2, 5],
[2, 5, 92],
[2, 5, 92, 688],
[2, 5, 92, 688, 58],
[2, 5, 92, 688, 58, 144],
[346, 29],
[346, 29, 6],
[346, 29, 6, 108],
[346, 29, 6, 108, 568],
[346, 29, 6, 108, 568, 884],
[408, 14],
[408, 14, 7],
[408, 14, 7, 885],
[408, 14, 7, 885, 1896],
[346, 64],
[346, 64, 380],
[346, 64, 380, 37],
[346, 64, 380, 37, 6],
[346, 64, 380, 37, 6, 47],
[346, 64, 380, 37, 6, 47, 689],
[346, 64, 380, 37, 6, 47, 689, 120],
[346, 64, 380, 37, 6, 47, 689, 120, 2],
[347, 1],
[347, 1, 1897],
[347, 1, 1897, 4],
[347, 1, 1897, 4, 8],
[347, 1, 1897, 4, 8, 257],
[347, 1, 1897, 4, 8, 257, 492],
[347, 1, 1897, 4, 8, 257, 492, 68],
[347, 1, 1897, 4, 8, 257, 492, 68, 87],
[347, 1, 1897, 4, 8, 257, 492, 68, 87, 149],
[57, 120],
[57, 120, 2],
[57, 120, 2, 347],
[408, 5],
[408, 5, 117],
[408, 5, 117, 5],
[408, 5, 117, 5, 139],
[408, 5, 117, 5, 139, 68],
[408, 5, 117, 5, 139, 68, 247],
[408, 5, 117, 5, 139, 68, 247, 1182],
```

```
[408, 5, 117, 5, 139, 68, 247, 1182, 63],
[48, 205],
[48, 205, 3],
[48, 205, 3, 16],
[48, 205, 3, 16, 409],
[140, 2],
[140, 2, 1898],
[140, 2, 1898, 348],
[140, 2, 1898, 348, 3],
[140, 2, 1898, 348, 3, 1],
[140, 2, 1898, 348, 3, 1, 493],
[408, 79],
[408, 79, 6],
[408, 79, 6, 46],
[408, 79, 6, 46, 124],
[140, 125],
[140, 125, 1899],
[140, 125, 1899, 322],
[140, 125, 1899, 322, 1183],
[140, 125, 1899, 322, 1183, 121],
[140, 125, 1899, 322, 1183, 121, 83],
[140, 125, 1899, 322, 1183, 121, 83, 1900],
[140, 125, 1899, 322, 1183, 121, 83, 1900, 6],
[1901, 407],
[1901, 407, 258],
[1901, 407, 258, 8],
[1901, 407, 258, 8, 494],
[1901, 407, 258, 8, 494, 79],
[1901, 407, 258, 8, 494, 79, 6],
[1901, 407, 258, 8, 494, 79, 6, 380],
[218, 408],
[140, 1902],
[140, 1902, 407],
[451, 94],
[451, 94, 24],
[451, 94, 24, 13],
[451, 94, 24, 13, 120],
[451, 94, 24, 13, 120, 63],
[48, 7],
[48, 7, 495],
[48, 7, 495, 4],
[48, 7, 495, 4, 28],
[451, 236],
[451, 236, 120],
[451, 236, 120, 236],
[451, 236, 120, 236, 46],
[451, 236, 120, 236, 46, 347],
```

```
[140, 24],
[140, 24, 258],
[140, 24, 258, 16],
[140, 24, 258, 16, 167],
[140, 24, 258, 16, 167, 886],
[140, 24, 258, 16, 167, 886, 131],
[140, 24, 258, 16, 167, 886, 131, 3],
[140, 24, 258, 16, 167, 886, 131, 3, 124],
[451, 5],
[451, 5, 29],
[451, 5, 29, 189],
[451, 5, 29, 189, 153],
[140, 120],
[140, 120, 887],
[140, 120, 887, 71],
[140, 120, 887, 71, 19],
[140, 120, 887, 71, 19, 33],
[140, 120, 887, 71, 19, 33, 1184],
[2, 31],
[2, 31, 14],
[2, 31, 14, 50],
[2, 31, 14, 50, 1903],
[2, 31, 14, 50, 1903, 136],
[2, 31, 14, 50, 1903, 136, 168],
[2, 31, 14, 50, 1903, 136, 168, 4],
[2, 31, 14, 50, 1903, 136, 168, 4, 28],
[888, 16],
[888, 16, 1904],
[888, 16, 1904, 410],
[888, 16, 1904, 410, 690],
[888, 16, 1904, 410, 690, 189],
[888, 16, 1904, 410, 690, 189, 4],
[888, 16, 1904, 410, 690, 189, 4, 72],
[259, 5],
[259, 5, 29],
[259, 5, 29, 1905],
[259, 5, 29, 1905, 28],
[259, 5, 29, 1905, 28, 889],
[17, 72],
[17, 72, 3],
[17, 72, 3, 257],
[17, 72, 3, 257, 1],
[17, 72, 3, 257, 1, 1906],
[17, 72, 3, 257, 1, 1906, 4],
[17, 72, 3, 257, 1, 1906, 4, 16],
[17, 72, 3, 257, 1, 1906, 4, 16, 124],
[11, 37],
```

```
[11, 37, 131],
[11, 37, 131, 16],
[11, 37, 131, 16, 1185],
[11, 37, 131, 16, 1185, 43],
[25, 70],
[25, 70, 1907],
[25, 70, 1907, 33],
[25, 70, 1907, 33, 190],
[25, 70, 1907, 33, 190, 2],
[25, 70, 1907, 33, 190, 2, 85],
[25, 70, 1907, 33, 190, 2, 85, 3],
[25, 70, 1907, 33, 190, 2, 85, 3, 9],
[48, 1186],
[48, 1186, 1186],
[48, 1186, 1186, 691],
[48, 1186, 1186, 691, 14],
[48, 1186, 1186, 691, 14, 890],
[451, 411],
[451, 411, 150],
[451, 411, 150, 7],
[451, 411, 150, 7, 237],
[2, 50],
[2, 50, 72],
[2, 50, 72, 206],
[2, 50, 72, 206, 131],
[2, 50, 72, 206, 131, 1908],
[2, 50, 72, 206, 131, 1908, 18],
[2, 50, 72, 206, 131, 1908, 18, 381],
[11, 36],
[11, 36, 27],
[11, 36, 27, 1909],
[11, 36, 27, 1909, 191],
[11, 36, 27, 1909, 191, 33],
[11, 36, 27, 1909, 191, 33, 1187],
[24, 34],
[24, 34, 177],
[24, 34, 177, 692],
[24, 34, 177, 692, 29],
[24, 34, 177, 692, 29, 189],
[48, 64],
[48, 64, 411],
[48, 64, 411, 34],
[48, 64, 411, 34, 150],
[2, 50],
[2, 50, 72],
[2, 50, 72, 139],
[2, 50, 72, 139, 407],
```

```
[2, 50, 72, 139, 407, 85],
[2, 50, 72, 139, 407, 85, 4],
[2, 50, 72, 139, 407, 85, 4, 16],
[346, 274],
[346, 274, 124],
[346, 274, 124, 4],
[346, 274, 124, 4, 38],
[95, 1910],
[95, 1910, 275],
[95, 1910, 275, 693],
[95, 1910, 275, 693, 249],
[95, 1910, 275, 693, 249, 1911],
[95, 1910, 275, 693, 249, 1911, 56],
[95, 1910, 275, 693, 249, 1911, 56, 1],
[95, 1910, 275, 693, 249, 1911, 56, 1, 1188],
[108, 147],
[108, 147, 15],
[108, 147, 15, 569],
[108, 147, 15, 569, 694],
[108, 147, 15, 569, 694, 1912],
[108, 147, 15, 569, 694, 1912, 11],
[108, 147, 15, 569, 694, 1912, 11, 219],
[108, 147, 15, 569, 694, 1912, 11, 219, 4],
[108, 147, 15, 569, 694, 1912, 11, 219, 4, 109],
[89, 51],
[89, 51, 9],
[89, 51, 9, 1189],
[89, 51, 9, 1189, 347],
[89, 51, 9, 1189, 347, 2],
[89, 51, 9, 1189, 347, 2, 8],
[89, 51, 9, 1189, 347, 2, 8, 66],
[1, 1190],
[1, 1190, 39],
[1, 1190, 39, 891],
[1, 1190, 39, 891, 103],
[140, 349],
[140, 349, 382],
[140, 349, 382, 80],
[140, 349, 382, 80, 4],
[57, 1],
[57, 1, 183],
[141, 89],
[141, 89, 9],
[141, 89, 9, 173],
[141, 89, 9, 173, 131],
[346, 10],
[346, 10, 1],
```

```
[346, 10, 1, 275],
[346, 10, 1, 275, 695],
[346, 10, 1, 275, 695, 61],
[346, 10, 1, 275, 695, 61, 1],
[346, 10, 1, 275, 695, 61, 1, 30],
[346, 10, 1, 275, 695, 61, 1, 30, 249],
[346, 10, 1, 275, 695, 61, 1, 30, 249, 145],
[140, 42],
[140, 42, 238],
[140, 42, 238, 7],
[140, 42, 238, 7, 1913],
[140, 42, 238, 7, 1913, 85],
[140, 42, 238, 7, 1913, 85, 3],
[140, 42, 238, 7, 1913, 85, 3, 9],
[140, 42, 238, 7, 1913, 85, 3, 9, 120],
[346, 452],
[346, 452, 9],
[346, 452, 9, 14],
[346, 452, 9, 14, 61],
[346, 452, 9, 14, 61, 1],
[346, 452, 9, 14, 61, 1, 30],
[346, 452, 9, 14, 61, 1, 30, 276],
[346, 452, 9, 14, 61, 1, 30, 276, 9],
[346, 452, 9, 14, 61, 1, 30, 276, 9, 120],
[383, 62],
[383, 62, 61],
[383, 62, 61, 9],
[383, 62, 61, 9, 1914],
[383, 62, 61, 9, 1914, 21],
[383, 62, 61, 9, 1914, 21, 17],
[383, 62, 61, 9, 1914, 21, 17, 1915],
[383, 62, 61, 9, 1914, 21, 17, 1915, 696],
[346, 9],
[346, 9, 67],
[346, 9, 67, 26],
[346, 9, 67, 26, 1191],
[346, 9, 67, 26, 1191, 90],
[140, 260],
[140, 260, 9],
[140, 260, 9, 120],
[48, 24],
[48, 24, 238],
[48, 24, 238, 42],
[48, 24, 238, 42, 11],
[48, 24, 238, 42, 11, 1916],
[48, 24, 238, 42, 11, 1916, 16],
[48, 24, 238, 42, 11, 1916, 16, 110],
```

```
[48, 24, 238, 42, 11, 1916, 16, 110, 4],
[48, 24, 238, 42, 11, 1916, 16, 110, 4, 124],
[350, 17],
[350, 17, 11],
[350, 17, 11, 192],
[350, 17, 11, 192, 2],
[350, 17, 11, 192, 2, 697],
[350, 17, 11, 192, 2, 697, 301],
[10, 75],
[10, 75, 1],
[10, 75, 1, 412],
[10, 75, 1, 412, 4],
[10, 75, 1, 412, 4, 496],
[10, 75, 1, 412, 4, 496, 193],
[76, 570],
[76, 570, 571],
[76, 570, 571, 41],
[76, 570, 571, 41, 109],
[76, 570, 571, 41, 109, 5],
[76, 570, 571, 41, 109, 5, 453],
[76, 570, 571, 41, 109, 5, 453, 80],
[76, 570, 571, 41, 109, 5, 453, 80, 85],
[140, 9],
[140, 9, 13],
[140, 9, 13, 892],
[346, 105],
[346, 105, 9],
[346, 105, 9, 1917],
[346, 105, 9, 1917, 154],
[48, 323],
[48, 323, 85],
[48, 323, 85, 85],
[48, 323, 85, 85, 5],
[48, 323, 85, 85, 5, 453],
[48, 323, 85, 85, 5, 453, 80],
[48, 323, 85, 85, 5, 453, 80, 85],
[218, 1],
[218, 1, 183],
[140, 71],
[140, 71, 277],
[140, 71, 277, 2],
[140, 71, 277, 2, 31],
[140, 71, 277, 2, 31, 14],
[140, 71, 277, 2, 31, 14, 377],
[346, 53],
[346, 53, 51],
[346, 53, 51, 120],
```

```
[346, 53, 51, 120, 6],
[346, 53, 51, 120, 6, 1192],
[346, 53, 51, 120, 6, 1192, 572],
[346, 53, 51, 120, 6, 1192, 572, 413],
[13, 14],
[13, 14, 16],
[13, 14, 16, 302],
[13, 14, 16, 302, 54],
[13, 14, 16, 302, 54, 39],
[13, 14, 16, 302, 54, 39, 1184],
[24, 117],
[24, 117, 6],
[24, 117, 6, 414],
[48, 225],
[48, 225, 8],
[48, 225, 8, 163],
[48, 225, 8, 163, 5],
[48, 225, 8, 163, 5, 146],
[48, 225, 8, 163, 5, 146, 14],
[48, 225, 8, 163, 5, 146, 14, 16],
[48, 225, 8, 163, 5, 146, 14, 16, 226],
[278, 1],
[278, 1, 1193],
[278, 1, 1193, 2],
[278, 1, 1193, 2, 178],
[278, 1, 1193, 2, 178, 1918],
[4, 111],
[4, 111, 142],
[4, 111, 142, 190],
[140, 13],
[140, 13, 9],
[140, 13, 9, 14],
[140, 13, 9, 14, 61],
[140, 13, 9, 14, 61, 1],
[140, 13, 9, 14, 61, 1, 30],
[48, 23],
[48, 23, 42],
[48, 23, 42, 238],
[48, 23, 42, 238, 3],
[48, 23, 42, 238, 3, 52],
[48, 23, 42, 238, 3, 52, 66],
[93, 59],
[93, 59, 1],
[93, 59, 1, 99],
[93, 59, 1, 99, 1194],
[93, 59, 1, 99, 1194, 25],
[93, 59, 1, 99, 1194, 25, 108],
```

```
[93, 59, 1, 99, 1194, 25, 108, 35],
[95, 134],
[95, 134, 893],
[95, 134, 893, 573],
[95, 134, 893, 573, 1919],
[27, 1920],
[27, 1920, 25],
[27, 1920, 25, 206],
[27, 1920, 25, 206, 95],
[27, 1920, 25, 206, 95, 10],
[27, 1920, 25, 206, 95, 10, 91],
[27, 1920, 25, 206, 95, 10, 91, 1921],
[27, 1920, 25, 206, 95, 10, 91, 1921, 1922],
[25, 1923],
[25, 1923, 1],
[25, 1923, 1, 1924],
[25, 1923, 1, 1924, 1925],
[25, 1923, 1, 1924, 1925, 35],
[25, 1923, 1, 1924, 1925, 35, 1],
[25, 1923, 1, 1924, 1925, 35, 1, 1195],
[71, 384],
[140, 119],
[140, 119, 690],
[140, 119, 690, 225],
[140, 119, 690, 225, 2],
[140, 119, 690, 225, 2, 894],
[140, 119, 690, 225, 2, 894, 58],
[140, 119, 690, 225, 2, 894, 58, 16],
[140, 119, 690, 225, 2, 894, 58, 16, 145],
[140, 119, 690, 225, 2, 894, 58, 16, 145, 566],
[17, 1926],
[17, 1926, 1927],
[17, 1926, 1927, 83],
[17, 1926, 1927, 83, 25],
[17, 1926, 1927, 83, 25, 277],
[17, 1926, 1927, 83, 25, 277, 41],
[17, 1926, 1927, 83, 25, 277, 41, 33],
[17, 1926, 1927, 83, 25, 277, 41, 33, 257],
[48, 10],
[48, 10, 24],
[48, 10, 24, 574],
[48, 10, 24, 574, 261],
[48, 10, 24, 574, 261, 3],
[48, 10, 24, 574, 261, 3, 1928],
[48, 10, 24, 574, 261, 3, 1928, 5],
[48, 10, 24, 574, 261, 3, 1928, 5, 65],
[48, 10, 24, 574, 261, 3, 1928, 5, 65, 14],
```

```
[19, 10],
[19, 10, 1],
[19, 10, 1, 1196],
[19, 10, 1, 1196, 2],
[19, 10, 1, 1196, 2, 1197],
[19, 10, 1, 1196, 2, 1197, 4],
[19, 10, 1, 1196, 2, 1197, 4, 8],
[19, 10, 1, 1196, 2, 1197, 4, 8, 1929],
[16, 1930],
[16, 1930, 106],
[16, 1930, 106, 384],
[16, 1930, 106, 384, 1931],
[16, 1930, 106, 384, 1931, 3],
[16, 1930, 106, 384, 1931, 3, 33],
[16, 1930, 106, 384, 1931, 3, 33, 279],
[140, 46],
[140, 46, 51],
[140, 46, 51, 411],
[140, 46, 51, 411, 150],
[140, 46, 51, 411, 150, 115],
[140, 46, 51, 411, 150, 115, 21],
[140, 46, 51, 411, 150, 115, 21, 25],
[140, 46, 51, 411, 150, 115, 21, 25, 11],
[140, 46, 51, 411, 150, 115, 21, 25, 11, 895],
[77, 16],
[77, 16, 275],
[77, 16, 275, 1932],
[77, 16, 275, 1932, 2],
[77, 16, 275, 1932, 2, 62],
[77, 16, 275, 1932, 2, 62, 1933],
[77, 16, 275, 1932, 2, 62, 1933, 257],
[27, 1198],
[27, 1198, 1934],
[27, 1198, 1934, 1],
[27, 1198, 1934, 1, 896],
[27, 1198, 1934, 1, 896, 4],
[27, 1198, 1934, 1, 896, 4, 1],
[27, 1198, 1934, 1, 896, 4, 1, 575],
[2, 77],
[2, 77, 93],
[2, 77, 93, 1935],
[2, 77, 93, 1935, 497],
[2, 77, 93, 1935, 497, 4],
[2, 77, 93, 1935, 497, 4, 1936],
[2, 77, 93, 1935, 497, 4, 1936, 698],
[2, 1937],
[2, 1937, 1938],
```

```
[2, 1937, 1938, 20],
[2, 1937, 1938, 20, 1939],
[2, 1937, 1938, 20, 1939, 4],
[2, 1937, 1938, 20, 1939, 4, 1199],
[77, 93],
[77, 93, 1940],
[77, 93, 1940, 4],
[77, 93, 1940, 4, 1200],
[77, 93, 1940, 4, 1200, 1941],
[77, 93, 1940, 4, 1200, 1941, 155],
[77, 93, 1940, 4, 1200, 1941, 155, 897],
[77, 93, 1940, 4, 1200, 1941, 155, 897, 1942],
[576, 14],
[576, 14, 1943],
[576, 14, 1943, 1],
[576, 14, 1943, 1, 1944],
[576, 14, 1943, 1, 1944, 56],
[576, 14, 1943, 1, 1944, 56, 1],
[576, 14, 1943, 1, 1944, 56, 1, 1945],
[24, 146],
[24, 146, 26],
[24, 146, 26, 699],
[24, 146, 26, 699, 11],
[24, 146, 26, 699, 11, 16],
[24, 146, 26, 699, 11, 16, 1946],
[24, 146, 26, 699, 11, 16, 1946, 149],
[156, 87],
[156, 87, 1],
[156, 87, 1, 124],
[156, 87, 1, 124, 1201],
[156, 87, 1, 124, 1201, 1947],
[156, 87, 1, 124, 1201, 1947, 17],
[156, 87, 1, 124, 1201, 1947, 17, 1],
[156, 87, 1, 124, 1201, 1947, 17, 1, 169],
[121, 280],
[121, 280, 11],
[121, 280, 11, 132],
[121, 280, 11, 132, 1948],
[121, 280, 11, 132, 1948, 21],
[48, 11],
[48, 11, 132],
[48, 11, 132, 5],
[58, 898],
[58, 898, 1],
[58, 898, 1, 1949],
[58, 898, 1, 1949, 281],
[58, 898, 1, 1949, 281, 27],
```

```
[58, 898, 1, 1949, 281, 27, 33],
[58, 898, 1, 1949, 281, 27, 33, 274],
[58, 898, 1, 1949, 281, 27, 33, 274, 30],
[155, 700],
[155, 700, 184],
[155, 700, 184, 19],
[155, 700, 184, 19, 51],
[155, 700, 184, 19, 51, 886],
[155, 700, 184, 19, 51, 886, 3],
[155, 700, 184, 19, 51, 886, 3, 72],
[59, 23],
[59, 23, 6],
[59, 23, 6, 65],
[59, 23, 6, 65, 41],
[59, 23, 6, 65, 41, 282],
[59, 23, 6, 65, 41, 282, 4],
[59, 23, 6, 65, 41, 282, 4, 701],
[1950, 1951],
[1950, 1951, 35],
[1950, 1951, 35, 41],
[1950, 1951, 35, 41, 7],
[1950, 1951, 35, 41, 7, 62],
[1950, 1951, 35, 41, 7, 62, 1952],
[1950, 1951, 35, 41, 7, 62, 1952, 1953],
[1954, 3],
[1954, 3, 1],
[1954, 3, 1, 1955],
[1954, 3, 1, 1955, 10],
[1954, 3, 1, 1955, 10, 75],
[1954, 3, 1, 1955, 10, 75, 33],
[1954, 3, 1, 1955, 10, 75, 33, 899],
[1954, 3, 1, 1955, 10, 75, 33, 899, 45],
[20, 27],
[20, 27, 16],
[20, 27, 16, 1202],
[20, 27, 16, 1202, 4],
[20, 27, 16, 1202, 4, 33],
[20, 27, 16, 1202, 4, 33, 454],
[20, 27, 16, 1202, 4, 33, 454, 170],
[20, 27, 16, 1202, 4, 33, 454, 170, 1956],
[20, 27, 16, 1202, 4, 33, 454, 170, 1956, 28],
[76, 1957],
[76, 1957, 16],
[76, 1957, 16, 282],
[76, 1957, 16, 282, 121],
[76, 1957, 16, 282, 121, 41],
[76, 1957, 16, 282, 121, 41, 7],
```

```
[76, 1957, 16, 282, 121, 41, 7, 702],
[76, 1957, 16, 282, 121, 41, 7, 702, 1958],
[64, 1959],
[64, 1959, 41],
[64, 1959, 41, 455],
[64, 1959, 41, 455, 2],
[64, 1959, 41, 455, 2, 1960],
[76, 1961],
[76, 1961, 17],
[76, 1961, 17, 15],
[76, 1961, 17, 15, 137],
[76, 1961, 17, 15, 137, 38],
[76, 1961, 17, 15, 137, 38, 194],
[76, 1961, 17, 15, 137, 38, 194, 15],
[76, 1961, 17, 15, 137, 38, 194, 15, 703],
[75, 25],
[75, 25, 704],
[75, 25, 704, 1962],
[75, 25, 704, 1962, 35],
[75, 25, 704, 1962, 35, 3],
[75, 25, 704, 1962, 35, 3, 1],
[75, 25, 704, 1962, 35, 3, 1, 1963],
[191, 1],
[191, 1, 75],
[191, 1, 75, 7],
[191, 1, 75, 7, 1964],
[191, 1, 75, 7, 1964, 1965],
[59, 1966],
[59, 1966, 41],
[59, 1966, 41, 33],
[59, 1966, 41, 33, 30],
[59, 1966, 41, 33, 30, 75],
[59, 1966, 41, 33, 30, 75, 108],
[59, 1966, 41, 33, 30, 75, 108, 705],
[3, 1],
[3, 1, 1967],
[3, 1, 1967, 4],
[3, 1, 1967, 4, 282],
[108, 25],
[108, 25, 456],
[108, 25, 456, 1968],
[108, 25, 456, 1968, 23],
[108, 25, 456, 1968, 23, 41],
[108, 25, 456, 1968, 23, 41, 1],
[108, 25, 456, 1968, 23, 41, 1, 275],
[108, 25, 456, 1968, 23, 41, 1, 275, 1969],
[2, 1970],
```

```
[2, 1970, 4],
[2, 1970, 4, 1],
[2, 1970, 4, 1, 1971],
[2, 1970, 4, 1, 1971, 1972],
[15, 415],
[15, 415, 3],
[15, 415, 3, 45],
[15, 415, 3, 45, 51],
[15, 415, 3, 45, 51, 73],
[15, 415, 3, 45, 51, 73, 385],
[15, 415, 3, 45, 51, 73, 385, 282],
[4, 1973],
[4, 1973, 1203],
[4, 1973, 1203, 498],
[4, 1973, 1203, 498, 2],
[4, 1973, 1203, 498, 2, 351],
[83, 10],
[83, 10, 1],
[83, 10, 1, 1974],
[83, 10, 1, 1974, 4],
[83, 10, 1, 1974, 4, 701],
[83, 10, 1, 1974, 4, 701, 107],
[83, 10, 1, 1974, 4, 701, 107, 2],
[83, 10, 1, 1974, 4, 701, 107, 2, 63],
[1975, 128],
[1975, 128, 7],
[1975, 128, 7, 499],
[1975, 128, 7, 499, 4],
[1975, 128, 7, 499, 4, 1976],
[1975, 128, 7, 499, 4, 1976, 1977],
[20, 1978],
[20, 1978, 2],
[20, 1978, 2, 1204],
[20, 1978, 2, 1204, 3],
[20, 1978, 2, 1204, 3, 106],
[20, 1978, 2, 1204, 3, 106, 1205],
[11, 83],
[11, 83, 7],
[11, 83, 7, 1979],
[11, 83, 7, 1979, 324],
[11, 83, 7, 1979, 324, 75],
[11, 83, 7, 1979, 324, 75, 13],
[11, 83, 7, 1979, 324, 75, 13, 32],
[11, 83, 7, 1979, 324, 75, 13, 32, 157],
[2, 9],
[2, 9, 156],
[2, 9, 156, 64],
```

```
[2, 9, 156, 64, 890],
[2, 9, 156, 64, 890, 577],
[2, 9, 156, 64, 890, 577, 33],
[2, 9, 156, 64, 890, 577, 33, 279],
[19, 3],
[19, 3, 900],
[19, 3, 900, 4],
[19, 3, 900, 4, 72],
[19, 3, 900, 4, 72, 41],
[19, 3, 900, 4, 72, 41, 706],
[19, 3, 900, 4, 72, 41, 706, 207],
[2, 901],
[2, 901, 1980],
[2, 901, 1980, 194],
[2, 901, 1980, 194, 1981],
[2, 901, 1980, 194, 1981, 703],
[27, 41],
[27, 41, 15],
[27, 41, 15, 97],
[27, 41, 15, 97, 416],
[27, 41, 15, 97, 416, 2],
[27, 41, 15, 97, 416, 2, 16],
[27, 41, 15, 97, 416, 2, 16, 5],
[27, 41, 15, 97, 416, 2, 16, 5, 136],
[27, 41, 15, 97, 416, 2, 16, 5, 136, 9],
[13, 1],
[13, 1, 902],
[13, 1, 902, 707],
[13, 1, 902, 707, 4],
[13, 1, 902, 707, 4, 33],
[13, 1, 902, 707, 4, 33, 1982],
[1, 1206],
[1, 1206, 4],
[1, 1206, 4, 16],
[1, 1206, 4, 16, 33],
[1, 1206, 4, 16, 33, 257],
[1, 1206, 4, 16, 33, 257, 2],
[1, 1206, 4, 16, 33, 257, 2, 1],
[1, 1206, 4, 16, 33, 257, 2, 1, 1207],
[1, 1206, 4, 16, 33, 257, 2, 1, 1207, 164],
[4, 16],
[4, 16, 1208],
[4, 16, 1208, 149],
[4, 16, 1208, 149, 2],
[4, 16, 1208, 149, 2, 1983],
[4, 16, 1208, 149, 2, 1983, 10],
[4, 16, 1208, 149, 2, 1983, 10, 1],
```

```
[4, 16, 1208, 149, 2, 1983, 10, 1, 575],
[57, 183],
[57, 183, 131],
[19, 417],
[19, 417, 1984],
[19, 417, 1984, 903],
[19, 417, 1984, 903, 89],
[19, 417, 1984, 903, 89, 9],
[19, 417, 1984, 903, 89, 9, 173],
[19, 417, 1984, 903, 89, 9, 173, 131],
[81, 1209],
[81, 1209, 9],
[81, 1209, 9, 179],
[81, 1209, 9, 179, 9],
[81, 1209, 9, 179, 9, 1210],
[81, 1209, 9, 179, 9, 1210, 21],
[81, 1209, 9, 179, 9, 1210, 21, 323],
[81, 1209, 9, 179, 9, 1210, 21, 323, 1985],
[37, 42],
[37, 42, 149],
[37, 42, 149, 227],
[37, 42, 149, 227, 457],
[37, 42, 149, 227, 457, 44],
[37, 42, 149, 227, 457, 44, 283],
[37, 42, 149, 227, 457, 44, 283, 4],
[37, 42, 149, 227, 457, 44, 283, 4, 352],
[85, 3],
[85, 3, 21],
[85, 3, 21, 37],
[85, 3, 21, 37, 63],
[85, 3, 21, 37, 63, 26],
[85, 3, 21, 37, 63, 26, 227],
[85, 3, 21, 37, 63, 26, 227, 46],
[85, 3, 21, 37, 63, 26, 227, 46, 167],
[85, 3, 21, 37, 63, 26, 227, 46, 167, 3],
[85, 3, 21, 37, 63, 26, 227, 46, 167, 3, 26],
[85, 3, 21, 37, 63, 26, 227, 46, 167, 3, 26, 174],
[11, 70],
[11, 70, 3],
[11, 70, 3, 80],
[11, 70, 3, 80, 47],
[11, 70, 3, 80, 47, 708],
[11, 70, 3, 80, 47, 708, 2],
[11, 70, 3, 80, 47, 708, 2, 284],
[11, 70, 3, 80, 47, 708, 2, 284, 3],
[11, 70, 3, 80, 47, 708, 2, 284, 3, 21],
[11, 70, 3, 80, 47, 708, 2, 284, 3, 21, 578],
```

```
[11, 70, 3, 80, 47, 708, 2, 284, 3, 21, 578, 3],
[11, 70, 3, 80, 47, 708, 2, 284, 3, 21, 578, 3, 21],
[37, 42],
[37, 42, 238],
[37, 42, 238, 1986],
[37, 42, 238, 1986, 3],
[37, 42, 238, 1986, 3, 52],
[37, 42, 238, 1986, 3, 52, 1211],
[37, 42, 238, 1986, 3, 52, 1211, 1212],
[75, 1213],
[75, 1213, 1987],
[75, 1213, 1987, 70],
[75, 1213, 1987, 70, 1214],
[75, 1213, 1987, 70, 1214, 60],
[75, 1213, 1987, 70, 1214, 60, 85],
[44, 37],
[44, 37, 42],
[44, 37, 42, 149],
[44, 37, 42, 149, 128],
[44, 37, 42, 149, 128, 1988],
[44, 37, 42, 149, 128, 1988, 10],
[44, 37, 42, 149, 128, 1988, 10, 52],
[44, 37, 42, 149, 128, 1988, 10, 52, 137],
[1989, 709],
[1989, 709, 10],
[1989, 709, 10, 1],
[1989, 709, 10, 1, 1990],
[1989, 709, 10, 1, 1990, 4],
[1989, 709, 10, 1, 1990, 4, 185],
[20, 75],
[20, 75, 55],
[20, 75, 55, 94],
[20, 75, 55, 94, 6],
[20, 75, 55, 94, 6, 579],
[20, 75, 55, 94, 6, 579, 353],
[20, 75, 55, 94, 6, 579, 353, 386],
[20, 75, 55, 94, 6, 579, 353, 386, 10],
[20, 75, 55, 94, 6, 579, 353, 386, 10, 129],
[85, 4],
[85, 4, 9],
[85, 4, 9, 323],
[85, 4, 9, 323, 2],
[85, 4, 9, 323, 2, 85],
[85, 4, 9, 323, 2, 85, 710],
[85, 4, 9, 323, 2, 85, 710, 9],
[85, 4, 9, 323, 2, 85, 710, 9, 347],
[140, 40],
```

```
[140, 40, 5],
[140, 40, 5, 1215],
[140, 40, 5, 1215, 58],
[140, 40, 5, 1215, 58, 9],
[140, 40, 5, 1215, 58, 9, 17],
[140, 40, 5, 1215, 58, 9, 17, 8],
[140, 40, 5, 1215, 58, 9, 17, 8, 1991],
[48, 47],
[48, 47, 37],
[48, 47, 37, 9],
[48, 47, 37, 9, 31],
[48, 47, 37, 9, 31, 14],
[48, 47, 37, 9, 31, 14, 247],
[346, 71],
[346, 71, 107],
[48, 71],
[48, 71, 107],
[140, 71],
[140, 71, 277],
[218, 183],
[34, 47],
[34, 47, 9],
[34, 47, 9, 500],
[34, 47, 9, 500, 325],
[34, 47, 9, 500, 325, 27],
[34, 47, 9, 500, 325, 27, 1216],
[3, 904],
[3, 904, 9],
[3, 904, 9, 1],
[3, 904, 9, 1, 208],
[3, 904, 9, 1, 208, 4],
[3, 904, 9, 1, 208, 4, 711],
[20, 9],
[20, 9, 13],
[20, 9, 13, 23],
[20, 9, 13, 23, 1],
[20, 9, 13, 23, 1, 303],
[20, 9, 13, 23, 1, 303, 1992],
[2, 33],
[2, 33, 1993],
[2, 33, 1993, 1217],
[2, 33, 1993, 1217, 1218],
[2, 33, 1993, 1217, 1218, 1994],
[346, 9],
[346, 9, 59],
[346, 9, 59, 209],
[346, 9, 59, 209, 3],
```

```
[346, 9, 59, 209, 3, 85],
[346, 9, 59, 209, 3, 85, 95],
[346, 9, 59, 209, 3, 85, 95, 1],
[346, 9, 59, 209, 3, 85, 95, 1, 580],
[346, 9, 59, 209, 3, 85, 95, 1, 580, 1219],
[48, 2],
[48, 2, 39],
[48, 2, 39, 9],
[48, 2, 39, 9, 1995],
[48, 2, 39, 9, 1995, 61],
[48, 2, 39, 9, 1995, 61, 7],
[48, 2, 39, 9, 1995, 61, 7, 581],
[48, 2, 39, 9, 1995, 61, 7, 581, 167],
[96, 7],
[96, 7, 1996],
[96, 7, 1996, 1997],
[96, 7, 1996, 1997, 5],
[96, 7, 1996, 1997, 5, 29],
[96, 7, 1996, 1997, 5, 29, 262],
[1, 580],
[1, 580, 11],
[1, 580, 11, 13],
[1, 580, 11, 13, 1],
[1, 580, 11, 13, 1, 712],
[1, 580, 11, 13, 1, 712, 3],
[1, 580, 11, 13, 1, 712, 3, 1],
[1, 580, 11, 13, 1, 712, 3, 1, 169],
[156, 17],
[156, 17, 15],
[156, 17, 15, 1998],
[156, 17, 15, 1998, 2],
[156, 17, 15, 1998, 2, 1999],
[156, 17, 15, 1998, 2, 1999, 2000],
[156, 17, 15, 1998, 2, 1999, 2000, 1220],
[1221, 1],
[1221, 1, 163],
[1221, 1, 163, 4],
[1221, 1, 163, 4, 169],
[1221, 1, 163, 4, 169, 2],
[1221, 1, 163, 4, 169, 2, 58],
[1221, 1, 163, 4, 169, 2, 58, 15],
[1221, 1, 163, 4, 169, 2, 58, 15, 2001],
[713, 10],
[713, 10, 501],
[713, 10, 501, 44],
[713, 10, 501, 44, 326],
[713, 10, 501, 44, 326, 10],
```

```
[713, 10, 501, 44, 326, 10, 185],
[713, 10, 501, 44, 326, 10, 185, 44],
[713, 10, 501, 44, 326, 10, 185, 44, 303],
[134, 2002],
[134, 2002, 2],
[134, 2002, 2, 2003],
[134, 2002, 2, 2003, 263],
[134, 2002, 2, 2003, 263, 2004],
[3, 15],
[3, 15, 1222],
[3, 15, 1222, 2],
[3, 15, 1222, 2, 4],
[3, 15, 1222, 2, 4, 1],
[3, 15, 1222, 2, 4, 1, 502],
[3, 15, 1222, 2, 4, 1, 502, 714],
[16, 905],
[16, 905, 2005],
[16, 905, 2005, 147],
[16, 905, 2005, 147, 2006],
[140, 9],
[140, 9, 2007],
[140, 9, 2007, 35],
[140, 9, 2007, 35, 1],
[140, 9, 2007, 35, 1, 2008],
[140, 9, 2007, 35, 1, 2008, 4],
[140, 9, 2007, 35, 1, 2008, 4, 1],
[140, 9, 2007, 35, 1, 2008, 4, 1, 580],
[106, 503],
[106, 503, 11],
[106, 503, 11, 285],
[106, 503, 11, 285, 582],
[106, 503, 11, 285, 582, 11],
[106, 503, 11, 285, 582, 11, 504],
[106, 503, 11, 285, 582, 11, 504, 173],
[715, 33],
[715, 33, 2009],
[715, 33, 2009, 2010],
[715, 33, 2009, 2010, 13],
[715, 33, 2009, 2010, 13, 2011],
[1, 1223],
[1, 1223, 4],
[1, 1223, 4, 2012],
[1, 1223, 4, 2012, 2013],
[1, 1223, 4, 2012, 2013, 38],
[1, 1223, 4, 2012, 2013, 38, 124],
[1, 1223, 4, 2012, 2013, 38, 124, 224],
[2, 39],
```

```
 [2, 39, 55],
 [2, 39, 55, 94],
 [2, 39, 55, 94, 32],
 [2, 39, 55, 94, 32, 263],
 [2, 39, 55, 94, 32, 263, 132],
 [2, 39, 55, 94, 32, 263, 132, 386],
 [2, 39, 55, 94, 32, 263, 132, 386, 2014],
 [1, 692],
 [1, 692, 36],
 [1, 692, 36, 583],
 [1, 692, 36, 583, 39],
 [1, 692, 36, 583, 39, 32],
 [1, 692, 36, 583, 39, 32, 2015],
 …]
```

[51]:
```
## Pad Sequences
max_sequence_len = max([len(x) for x in input_sequences])
max_sequence_len
```

[51]: 14

[52]:
```
input_sequences = np.array(pad_sequences(input_sequences,␣
 ↪maxlen=max_sequence_len, padding='pre'))
input_sequences
```

[52]:
```
array([[   0,    0,    0, …,    0,    1,  687],
       [   0,    0,    0, …,    1,  687,    4],
       [   0,    0,    0, …,  687,    4,   45],
       …,
       [   0,    0,    0, …,    4,   45, 1047],
       [   0,    0,    0, …,   45, 1047,    4],
       [   0,    0,    0, …, 1047,    4,  193]], dtype=int32)
```

[53]:
```
##create predicitors and label
import tensorflow as tf
x, y = input_sequences[:,:-1], input_sequences[:,-1]
```

[54]:
```
x
```

[54]:
```
array([[   0,    0,    0, …,    0,    0,    1],
       [   0,    0,    0, …,    0,    1,  687],
       [   0,    0,    0, …,    1,  687,    4],
       …,
       [   0,    0,    0, …,  687,    4,   45],
       [   0,    0,    0, …,    4,   45, 1047],
       [   0,    0,    0, …,   45, 1047,    4]], dtype=int32)
```

[55]:
```
y
```

[55]:
```
array([ 687,    4,   45, …, 1047,    4,  193], dtype=int32)
```

[56]:
```
y = tf.keras.utils.to_categorical(y, num_classes=total_words)
y
```

[56]:
```
array([[0., 0., 0., …, 0., 0., 0.],
       [0., 0., 0., …, 0., 0., 0.],
       [0., 0., 0., …, 0., 0., 0.],
       …,
       [0., 0., 0., …, 0., 0., 0.],
       [0., 0., 0., …, 0., 0., 0.],
       [0., 0., 0., …, 0., 0., 0.]])
```

[57]:
```
# Split the data into training and testing sets
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2)
```

[73]:
```
# Define early stopping
from tensorflow.keras.callbacks import EarlyStopping
early_stopping = EarlyStopping(monitor='val_loss', patience=10,␣
 ↪restore_best_weights=True)
```

[74]:
```
# Define model
model = Sequential()
model.add(Embedding(total_words,100,input_length=max_sequence_len-1))
model.add(LSTM(150, return_sequences=True))
model.add(Dropout(0.2))
model.add(LSTM(100))
model.add(Dense(total_words, activation='softmax'))

# Manually build the model
model.build(input_shape=(None, max_sequence_len))
model.
 ↪compile(loss="categorical_crossentropy",optimizer='adam',metrics=['accuracy'])

# Summary
model.summary()
```

```
Model: "sequential_14"
```

Layer (type)	Output Shape	Param #
embedding_14 (Embedding)	(None, 14, 100)	481,800
lstm_26 (LSTM)	(None, 14, 150)	150,600

```
dropout_13 (Dropout)            (None, 14, 150)                0

lstm_27 (LSTM)                  (None, 100)                100,400

dense_13 (Dense)                (None, 4818)               486,618
```

Total params: 1,219,418 (4.65 MB)

Trainable params: 1,219,418 (4.65 MB)

Non-trainable params: 0 (0.00 B)

[75]:
```
## Train the model
history=model.
  ↪fit(x_train,y_train,epochs=50,validation_data=(x_test,y_test),verbose=1,callbacks=[early_st
```

```
Epoch 1/50
644/644             44s 63ms/step -
accuracy: 0.0311 - loss: 7.1298 - val_accuracy: 0.0332 - val_loss: 6.7540
Epoch 2/50
644/644             45s 70ms/step -
accuracy: 0.0368 - loss: 6.4326 - val_accuracy: 0.0431 - val_loss: 6.8249
Epoch 3/50
644/644             76s 61ms/step -
accuracy: 0.0412 - loss: 6.3238 - val_accuracy: 0.0499 - val_loss: 6.9043
Epoch 4/50
644/644             43s 63ms/step -
accuracy: 0.0486 - loss: 6.1749 - val_accuracy: 0.0499 - val_loss: 6.9207
Epoch 5/50
644/644             39s 60ms/step -
accuracy: 0.0544 - loss: 6.0428 - val_accuracy: 0.0558 - val_loss: 6.9800
Epoch 6/50
644/644             39s 61ms/step -
accuracy: 0.0598 - loss: 5.9079 - val_accuracy: 0.0606 - val_loss: 7.0251
Epoch 7/50
644/644             43s 66ms/step -
accuracy: 0.0691 - loss: 5.7336 - val_accuracy: 0.0643 - val_loss: 7.0696
Epoch 8/50
644/644             79s 63ms/step -
accuracy: 0.0829 - loss: 5.6246 - val_accuracy: 0.0651 - val_loss: 7.1536
Epoch 9/50
644/644             39s 61ms/step -
accuracy: 0.0852 - loss: 5.4773 - val_accuracy: 0.0676 - val_loss: 7.1941
```

```
Epoch 10/50
644/644             39s 61ms/step -
accuracy: 0.0930 - loss: 5.3455 - val_accuracy: 0.0694 - val_loss: 7.2945
Epoch 11/50
644/644             39s 61ms/step -
accuracy: 0.0978 - loss: 5.2414 - val_accuracy: 0.0701 - val_loss: 7.4178
```

[76]:
```
# Function to predict the next word
def predict_next_word(model, tokenizer, text, max_sequence_len):
    token_list = tokenizer.texts_to_sequences([text])[0]
    if len(token_list) >= max_sequence_len:
        token_list = token_list[-(max_sequence_len-1):]  # Ensure the sequence␣
 ↪length matches max_sequence_len-1
    token_list = pad_sequences([token_list], maxlen=max_sequence_len-1,␣
 ↪padding='pre')
    predicted = model.predict(token_list, verbose=0)
    predicted_word_index = np.argmax(predicted, axis=1)
    for word, index in tokenizer.word_index.items():
        if index == predicted_word_index:
            return word
    return None
```

[84]:
```
input_text="The words are worst"
print(f"Input text:{input_text}")
max_sequence_len=model.input_shape[1]+1
next_word=predict_next_word(model,tokenizer,input_text,max_sequence_len)
print(f"Next Word PRediction:{next_word}")
```

```
Input text:The words are worst
Next Word PRediction:the
```

[85]:
```
model.save('next_word_lst.h5')
import pickle
with open('tokenizer.pickle','wb') as handle:
    pickle.dump(tokenizer,handle,protocol=pickle.HIGHEST_PROTOCOL)
```

```
WARNING:absl:You are saving your model as an HDF5 file via `model.save()` or
`keras.saving.save_model(model)`. This file format is considered legacy. We
recommend using instead the native Keras format, e.g.
`model.save('my_model.keras')` or `keras.saving.save_model(model,
'my_model.keras')`.
```