In [148... import pandas as pd import nltk import numpv import sys from keras.models import Sequential from keras.layers import Dense from keras.layers import Dropout from keras.layers import LSTM from keras.callbacks import ModelCheckpoint from keras.utils import np_utils import string import numpy as np from tensorflow.keras.preprocessing.text import Tokenizer from tensorflow.keras.utils import to_categorical from tensorflow.keras.models import Sequential from keras.layers import Dense, LSTM, Embedding from tensorflow.keras.preprocessing.sequence import pad sequences Opening the nursery rhymes and separating them based on line In [1]: file = open('nursery_rhymes.txt','r') with open('nursery_rhymes.txt') as f: contents = f.read() data=contents.split('\n') In [150... len(data) 5071 Out[150]: In [151... data = " ".join(data) Cleaning data and separating by individual words In [152... def clean_text(doc): tokens = doc.split() #Removing punctuation table = str.maketrans('', '', string.punctuation) tokens = [w.translate(table) for w in tokens] #Removes non alphabet characters tokens = [word for word in tokens if word.isalpha()] #Makes everything lowercase tokens = [word.lower() for word in tokens] return tokens In [153... tokens = clean_text(data) print(tokens[:10]) ['the', 'queen', 'of', 'hearts', 'the', 'queen', 'of', 'hearts', 'she', 'made'] I can now see that the data has been tokenized and now it is easy to read with no random characters included with everything in a unliform character set In [154... | print("There are: "+ str(len(set(tokens)))+ " unique words") print("There are: "+ str(len(tokens))+ " total words") There are: 2444 unique words There are: 16786 total words In [155... length = 10 + 1 #Appending 10 words to a line that is fed in to create a sequence where the model will then determine what the following word # will be until it has finished the nursery Rhyme for i in range(length, len(tokens)): seq = tokens[i-length:i] line = ' '.join(seq) lines.append(line) **if** i>200000: break print(len(lines)) 16775 In [156... lines[3] 'hearts the queen of hearts she made some tarts all on' Assigning integer values to each word so it can be fed into the model In [157... tokenizer = Tokenizer() tokenizer.fit_on_texts(lines) sequences = tokenizer.texts_to_sequences(lines) In [158... sequences=np.array(sequences) X, y = sequences[:, :-1], sequences[:,-1]wordDictSize = len(tokenizer.word_index) + 1 y = to_categorical(y, num_classes=wordDictSize) sequence_length = X.shape[1] **Building model** In [160... print(sequence_length) model = Sequential() model.add(Embedding(wordDictSize, 50, input_length = sequence_length)) model.add(LSTM(100, return_sequences= True)) model.add(LSTM(100)) model.add(Dense(100, activation = 'relu')) model.add(Dense(wordDictSize, activation='softmax')) model.summary() 10 Model: "sequential_11" Layer (type) Output Shape Param # ______ embedding_8 (Embedding) (None, 10, 50) 122250 lstm_18 (LSTM) 60400 (None, 10, 100) lstm_19 (LSTM) (None, 100) 80400 dense_18 (Dense) (None, 100) 10100 dense_19 (Dense) 246945 (None, 2445) ______ Total params: 520,095 Trainable params: 520,095 Non-trainable params: 0 Compile and fit model.compile(loss = 'categorical crossentropy', optimizer = 'adam', metrics = ['accuracy']) In [163... $model.fit(X, y, batch_size=256, epochs = 300)$ Epoch 1/300 Epoch 2/300 Epoch 3/300 Epoch 4/300 Epoch 5/300 Epoch 6/300 4s 57ms/step - loss: 6.1898 - accuracy: 0.0548 66/66 [== Epoch 7/300 Epoch 8/300 Epoch 9/300 Epoch 10/300 Epoch 11/300 Epoch 12/300 Epoch 13/300 Epoch 14/300 Epoch 15/300 Epoch 16/300 Epoch 17/300 Epoch 18/300 Epoch 19/300 Epoch 20/300 Epoch 21/300 Epoch 22/300 Epoch 23/300 Epoch 24/300 Epoch 25/300 Epoch 26/300 Epoch 27/300 Epoch 28/300 Epoch 29/300 Epoch 30/300 Epoch 31/300 Epoch 32/300 Epoch 33/300 Epoch 34/300 Epoch 35/300 Epoch 36/300 Epoch 37/300 Epoch 38/300 Epoch 39/300 Epoch 40/300 Epoch 41/300 Epoch 42/300 Epoch 43/300 66/66 [===== Epoch 44/300 Epoch 45/300 Epoch 46/300 Epoch 47/300 Epoch 49/300 Epoch 50/300 Epoch 51/300 Epoch 52/300 Epoch 53/300 Epoch 54/300 Epoch 55/300 Epoch 56/300 Epoch 57/300 Epoch 58/300 Epoch 59/300 Epoch 60/300 Epoch 61/300 Epoch 62/300 Epoch 63/300 Epoch 64/300 Epoch 65/300 Epoch 66/300 Epoch 67/300 Epoch 68/300 Epoch 69/300 Epoch 70/300 Epoch 71/300 Epoch 72/300 Epoch 73/300 Epoch 74/300 Epoch 75/300 Epoch 76/300 Epoch 77/300 Epoch 78/300 Epoch 79/300 Epoch 80/300 Epoch 81/300 Epoch 82/300 Epoch 83/300 Epoch 84/300 Epoch 85/300 Epoch 86/300 Epoch 87/300 Epoch 88/300 Epoch 89/300 Epoch 90/300 Epoch 91/300 Epoch 92/300 Epoch 93/300 Epoch 94/300 Epoch 95/300 Epoch 96/300 Epoch 97/300 Epoch 98/300 Epoch 99/300 Epoch 100/300 Epoch 101/300 Epoch 102/300 Epoch 103/300 Epoch 104/300 Epoch 105/300 Epoch 106/300 Epoch 107/300 Epoch 108/300 Epoch 109/300 Epoch 110/300 Epoch 111/300 Epoch 112/300 Epoch 113/300 Epoch 114/300 Epoch 115/300 Epoch 116/300 Epoch 117/300 66/66 [===========================] - 4s 56ms/step - loss: 0.8755 - accuracy: 0.7873 Epoch 118/300 Epoch 119/300 Epoch 120/300 Epoch 121/300 Epoch 122/300 Epoch 123/300 Epoch 124/300 Epoch 125/300 Epoch 126/300 Epoch 127/300 Epoch 128/300 Epoch 129/300 Epoch 130/300 Epoch 131/300 66/66 [====================] - 4s 57ms/step - loss: 0.6901 - accuracy: 0.8331 Epoch 132/300 Epoch 133/300 Epoch 134/300 Epoch 135/300 Epoch 136/300 Epoch 137/300 Epoch 138/300 Epoch 139/300 66/66 [====================] - 4s 57ms/step - loss: 0.5123 - accuracy: 0.8852 Epoch 140/300 66/66 [===================] - 4s 57ms/step - loss: 0.5070 - accuracy: 0.8847 Epoch 141/300 Epoch 142/300 Epoch 143/300 Epoch 144/300 Epoch 145/300 Epoch 146/300 Epoch 147/300 Epoch 148/300 Epoch 149/300 66/66 [===================] - 4s 57ms/step - loss: 0.5068 - accuracy: 0.8788 Epoch 150/300 Epoch 151/300 Epoch 152/300 Epoch 153/300 Epoch 154/300 Epoch 155/300 Epoch 156/300 Epoch 157/300 Epoch 158/300 Epoch 159/300 Epoch 160/300 Epoch 161/300 Epoch 162/300 Epoch 163/300 66/66 [===================] - 4s 57ms/step - loss: 0.3153 - accuracy: 0.9344 Epoch 164/300 Epoch 165/300 Epoch 166/300 Epoch 167/300 Epoch 168/300 Epoch 169/300 Epoch 170/300 Epoch 171/300 Epoch 172/300 Epoch 173/300 Epoch 174/300 Epoch 175/300 Epoch 176/300 Epoch 177/300 Epoch 178/300 66/66 [====================] - 4s 58ms/step - loss: 0.1604 - accuracy: 0.9741 Epoch 179/300 Epoch 180/300 Epoch 181/300 Epoch 182/300 Epoch 183/300 66/66 [====================] - 4s 58ms/step - loss: 0.1662 - accuracy: 0.9731 Epoch 184/300 Epoch 185/300 Epoch 186/300 Epoch 187/300 Epoch 188/300 Epoch 189/300 66/66 [====================] - 4s 58ms/step - loss: 0.3706 - accuracy: 0.9063 Epoch 190/300 Epoch 191/300 Epoch 192/300 Epoch 193/300 Epoch 194/300 Epoch 195/300 Epoch 196/300 66/66 [====================] - 4s 58ms/step - loss: 0.0942 - accuracy: 0.9887 Epoch 197/300 66/66 [===================] - 4s 59ms/step - loss: 0.0904 - accuracy: 0.9886 Epoch 198/300 Epoch 199/300 Epoch 200/300 Epoch 201/300 Epoch 202/300 Epoch 203/300 Epoch 204/300 Epoch 205/300 Epoch 206/300 Epoch 207/300 Epoch 208/300 Epoch 209/300 Epoch 210/300 Epoch 211/300 Epoch 212/300 Epoch 213/300 66/66 [====================] - 4s 58ms/step - loss: 0.4202 - accuracy: 0.8821 Epoch 214/300 66/66 [===================] - 4s 59ms/step - loss: 0.2571 - accuracy: 0.9361 Epoch 215/300 66/66 [===================] - 4s 58ms/step - loss: 0.1696 - accuracy: 0.9641 Epoch 216/300 Epoch 217/300 Epoch 218/300 Epoch 219/300 Epoch 220/300 Epoch 221/300 Epoch 222/300 66/66 [===================] - 4s 59ms/step - loss: 0.0519 - accuracy: 0.9937 Epoch 223/300 Epoch 224/300 Epoch 225/300 Epoch 226/300 Epoch 227/300 Epoch 228/300 Epoch 229/300 Epoch 230/300 Epoch 231/300 Epoch 232/300 Epoch 233/300 Epoch 234/300 Epoch 235/300 Epoch 236/300 Epoch 237/300 Epoch 238/300 Epoch 239/300 Epoch 240/300 Epoch 241/300 66/66 [===================] - 4s 58ms/step - loss: 0.9199 - accuracy: 0.7577 Epoch 242/300 Epoch 243/300 66/66 [====================] - 4s 58ms/step - loss: 0.3765 - accuracy: 0.8950 Epoch 244/300 66/66 [===================] - 4s 59ms/step - loss: 0.2050 - accuracy: 0.9493 Epoch 245/300 66/66 [===================] - 4s 59ms/step - loss: 0.1063 - accuracy: 0.9813 Epoch 246/300 Epoch 247/300 Epoch 248/300 Epoch 249/300 Epoch 250/300 Epoch 251/300 Epoch 252/300 Epoch 253/300 Epoch 254/300 Epoch 255/300 Epoch 256/300 Epoch 257/300 Epoch 258/300 Epoch 259/300 Epoch 260/300 Epoch 261/300 Epoch 262/300 66/66 [====================] - 4s 58ms/step - loss: 0.0303 - accuracy: 0.9944 Epoch 263/300 66/66 [====================] - 4s 59ms/step - loss: 0.0287 - accuracy: 0.9948 Epoch 264/300 Epoch 265/300 Epoch 266/300 Epoch 267/300 Epoch 268/300 Epoch 269/300 Epoch 270/300 Epoch 271/300 66/66 [===================] - 4s 59ms/step - loss: 0.0289 - accuracy: 0.9944 Epoch 272/300 Epoch 273/300 Epoch 274/300 Epoch 275/300 Epoch 276/300 66/66 [====================] - 4s 59ms/step - loss: 0.4964 - accuracy: 0.8631 Epoch 277/300 Epoch 278/300 Epoch 279/300 Epoch 280/300 Epoch 281/300 Epoch 282/300 Epoch 283/300 Epoch 284/300 Epoch 285/300 66/66 [==============] - 4s 60ms/step - loss: 0.0282 - accuracy: 0.9947 Epoch 286/300 Epoch 287/300 66/66 [===================] - 4s 59ms/step - loss: 0.0267 - accuracy: 0.9951 Epoch 288/300 Epoch 289/300 Epoch 290/300 Epoch 291/300 Epoch 292/300 Epoch 293/300 Epoch 294/300 Epoch 295/300 Epoch 296/300 Epoch 297/300 Epoch 298/300 Epoch 299/300 Epoch 300/300 <keras.callbacks.History at 0x17ad1d6ffa0> Out[163]: I am actually really surprised with this accuracy as it is extremely high. Creating the rhyme by generating text based on what the model learned In [194... def create_rhyme(model, tokenizer, text_seq_length, seed, num_words): rhyme=[] for _ in range(num_words): encoded = tokenizer.texts_to_sequences([seed])[0] encoded=pad_sequences([encoded], maxlen = text_seq_length, truncating='pre') predict_y=model.predict(encoded) y_pred=np.argmax(predict_y,axis=1) pred_word ='' for word, index in tokenizer.word_index.items(): if (index == y_pred): pred_word = word break seed = seed + " " + pred_word rhyme.append(pred_word) return ' '.join(rhyme) This creates a rhyme by trying to predict the next word continuously until the number of words it generates is complete # choosing random seed to make the nursery rhymes for import random seedLen=(len(lines)) seedIndex=random.randrange(seedLen) seed=lines[seedIndex] #Creating Rhyme raw_rhyme=create_rhyme(model, tokenizer, sequence_length, seed, 600) Formatting the rhyme to 30 lines with 20 words each In [196... rhyme=raw_rhyme.split() counter=0 temp="" print("Model Generated Nursery Rhyme: ") print(" ") for i in range(len(rhyme)): temp+=rhyme[i]+" " **if** (i+1)%20==0 and i!=0: counter+=1 print(temp) print("") temp="" Model Generated Nursery Rhyme: woman went fortunetelling by cherrystones one i love two i love three i love i say four i love with all my heart five i cast away six he loves seven she loves eight both love nine he comes ten he tarries eleven he courts and twelve he marries little bopeep little bopeep has lost her sheep and cant tell where to find them let them alone and theyll come home and bring their tails behind them little bopeep fell fast asleep and dreamt she heard them bleating and when she awoke she found it a joke for still they were all fleeting then up she took her little crook determined for to find them she found them indeed but it made her heart bleed for theyd left all their tails behind them it happened one day as bopeep did stray into a meadow hard by there she espied their tails side by side all hung on a tree to dry she heaved a sigh and wiped her eye and went over hill and dale oh and tried what she could as a shepherdess should to tack to each sheep its tail oh to bed come lets to bed says sleepyhead sit up a while says slow put on the pan says greedy nan lets sup before we go of going to bed go to bed first a golden purse go to bed second a golden pheasant go to bed third a golden bird grace before meat here a little child i stand heaving up my either hand cold as paddocks though they be here i lift them up to thee for a benison to fall on our meat and on us all there was a butcher there was a butcher cut his thumb when it did bleed then blood did come there was a chandler making candle when he them stript he did them handle there was a cobbler clouting shoon when they were mended they were done there was a crow sat on a stone when he was gone then there was none there was a horse going to the mill when he went on he stood not still there was a lackey ran a race when he ran fast he ran apace there was a monkey climbed a tree when he fell down then down fell he there was a navy went into spain when it returnd it came again there was an old woman lived under a hill and if shes not gone she lives there still winter has come cold and raw the north wind doth blow bleak in a morning early all the hills are covered with snow and winters now come fairly mondays child mondays child is fair of face tuesdays child is full of grace wednesdays child is full of woe thursdays child has far to go fridays child is loving and giving saturdays child works hard for its living but the child that is born on the sabbath day is bonny and blithe and good and gay jack and jill jack and jill went up the hill to fetch a pail of water jack fell down and broke his crown and jill came tumbling after then up jack got up and off did trot as fast as he could caper to old dame dob who patched his nob with vinegar and brown paper charley charley charley stole the barley out of the bakers shop the baker came out and gave him a clout which made poor charley hop the pipers cow there was a piper had a cow and he While some of the rhymes dont make any sense, I was actually surprised that there was some rhymeing scheme in parts of it considering this was not written by a human