

Scenario

Throughout your career as a data analyst, you will assess continuing data sources for their relevance to specific research questions. Organizations use data sets to analyze their operations. Organizations have many possible uses for these data sets to support their decision-making processes.

In your previous coursework, you have explored a variety of supervised and unsupervised data mining models. You have seen the power of using data analytical techniques to help organizations make data-driven decisions and now want to extend these models into areas of machine learning and artificial intelligence. In this course, you will explore the use of neural networks and natural language processing (NLP).

Data Selected <https://archive.ics.uci.edu/ml/datasets/Sentiment+Labelled+Sentences>

Part I: Research Question

A. Describe the purpose of this data analysis by doing the following:

A1. Summarize one research question that you will answer using neural network models and NLP techniques. Be sure the research question is relevant to a real-world organizational situation and sentiment analysis captured in your chosen dataset.

Rubric: The submission summarizes 1 research question, and the research question is relevant to a realistic organizational situation and sentiment analysis. The research question can be answered using the selected dataset, neural network model, and NLP techniques.

Is it possible to analyze customer sentiment using Neural Networks and the Natural Language Processing technique so that this company can take appropriate actions to improve customer satisfaction?

A2. Define the objectives or goals of the data analysis. Be sure the objectives or goals are reasonable within the scope of the research question and are represented in the available data.

Rubric: The submission clearly defines each of the objectives or goals of the data analysis. Each objective or goal is reasonable within the scope of the research question and represented in the available data.

The goal of this data analysis is to attempt to predict the customer sentiment (how a user feels) of a product or service based off the word choices that the customer has chosen.

A3. Identify a type of neural network capable of performing a text classification task that can be trained to produce useful predictions on text sequences on the selected data set.

Rubric: The submission identifies an industry-relevant type of neural network that can be trained to produce useful text classification predictions on text sequences on the selected dataset.

A Bidirectional Recurrent Neural Network (RNN) is capable of performing a sentiment analysis and text classification-related task including Sequence Labeling, Speech, Tagging, and more. Bidirectional RNNs maximize the order sensitivity of an RNN. An LSTM layer was input to the Bidirectional layer in order to catch the complexities that a single RNN might miss. In other words, Bidirectional RNNs have two layers where one of the layers interprets the sequences in chronological order and the other in anti-chronological order.

Part II: Data Preparation

B. Summarize the data cleaning process by doing the following:

B1. Perform exploratory data analysis on the chosen dataset, and include an explanation of each of the following elements:

Rubric: The submission performs exploratory data analysis on the chosen dataset and includes an explanation of each of the 4 listed elements. Each element is accurate and aligns to the chosen data set.

- presence of unusual characters (e.g., emojis, non-English characters, etc.)
- vocabulary size
- proposed word embedding length
- statistical justification for the chosen maximum sequence length

Data Set Chosen

Source: Generating Personalized Recipes from Historical User Preferences Bodhisattwa Prasad Majumder, *Shuyang Li*, Jianmo Ni, Julian McAuley EMNLP, 2019 pdf

```
In [1]: from platform import python_version
print("python version: ", python_version())
```

python version: 3.8.5

```
In [2]: # LOADING PYTHON LIBRARIES NEEDED FOR THE ANALYSIS
import sys
import os

#scikit-learn for predictive analysis
import sklearn
from sklearn import preprocessing
```

```

from sklearn.preprocessing import OneHotEncoder
from sklearn import model_selection
import re
import tensorflow as tf
import keras

import tensorflow as ft
from tensorflow import keras
from keras import preprocessing

from tensorflow.keras.preprocessing.text import Tokenizer # To convert input text to to
from tensorflow.keras.callbacks import EarlyStopping # To stop training at a threshold
from tensorflow.keras.models import load_model # To Load a saved model
from tensorflow.keras.layers import Activation, Dropout, Dense, Embedding, Conv1D, LSTM
from keras.layers import Flatten
from keras.layers import GlobalMaxPooling1D
from tensorflow.keras.models import Sequential
from tensorflow.keras.preprocessing.sequence import pad_sequences # For text padding

import pandas as pd # For dataframe and data manipulations
import numpy as np # Provides array objects for calculations
import gzip # For unzipping the input customer review files
import seaborn as sns # For visualization like data distribution matrix
import nltk # NLP toolkit
import re # For regular expression

from nltk.corpus import stopwords # For stopword usage
from nltk import word_tokenize
from nltk.stem import PorterStemmer
nltk.download('stopwords')
nltk.download('punkt')
nltk.download('wordnet')

import numpy
from scipy.signal import max_len_seq
import matplotlib.pyplot as plt
%matplotlib inline
sns.color_palette("vlag", as_cmap=True)

```

```

[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\verab\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[nltk_data] Downloading package punkt to
[nltk_data] C:\Users\verab\AppData\Roaming\nltk_data...
[nltk_data] Package punkt is already up-to-date!
[nltk_data] Downloading package wordnet to
[nltk_data] C:\Users\verab\AppData\Roaming\nltk_data...
[nltk_data] Package wordnet is already up-to-date!

```

Out[2]: <matplotlib.colors.ListedColormap at 0x245f6563160>

In [3]:

```

# Aesthetics
import warnings

warnings.filterwarnings('ignore')
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)

# plots for exploratory data analysis
sns.set_style('whitegrid')

```

```
In [4]: # Import the initial dataset into an intial_df
# Read the raw data

df_amazon = pd.read_csv('./sentiment labelled sentences/amazon_cells_labelled.txt', del
df_amazon.columns=['review','sentiment']

df_imdb = pd.read_csv('./sentiment labelled sentences/imdb_labelled.txt', delimiter='\t
df_imdb.columns=['review','sentiment']

df_yelp = pd.read_csv('./sentiment labelled sentences/yelp_labelled.txt', delimiter='\t
df_yelp.columns=['review','sentiment']
```

```
In [5]: # Look at the shapre of the data
print(df_amazon.shape)
print(df_imdb.shape)
print(df_yelp.shape)
```

```
(1000, 2)
(748, 2)
(1000, 2)
```

```
In [6]: df_amazon.describe()
```

```
Out[6]:
```

	sentiment
count	1000.00000
mean	0.50000
std	0.50025
min	0.00000
25%	0.00000
50%	0.50000
75%	1.00000
max	1.00000

```
In [7]: df_imdb.describe()
```

```
Out[7]:
```

	sentiment
count	748.000000
mean	0.516043
std	0.500077
min	0.000000
25%	0.000000
50%	1.000000

	sentiment
75%	1.000000
max	1.000000

In [8]:

```
df_yelp.describe()
```

Out[8]:

	sentiment
count	1000.00000
mean	0.50000
std	0.50025
min	0.00000
25%	0.00000
50%	0.50000
75%	1.00000
max	1.00000

In [9]:

```
frames = [df_amazon, df_imdb, df_yelp]
df = pd.concat(frames)
df.describe()
```

Out[9]:

	sentiment
count	2748.000000
mean	0.504367
std	0.500072
min	0.000000
25%	0.000000
50%	1.000000
75%	1.000000
max	1.000000

In [10]:

```
# presence of unusual characters (e.g., emojis, non-English characters, etc.)
pd.set_option('display.max_colwidth', 5000)
df = df_amazon
df.head(10)
```

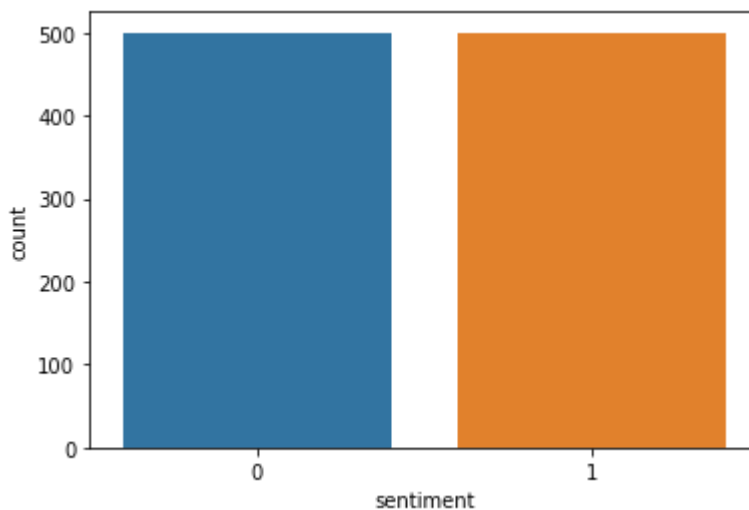
Out[10]:

	review	sentiment
0	So there is no way for me to plug it in here in the US unless I go by a converter.	0
1	Good case, Excellent value.	1

	review	sentiment
2	Great for the jawbone.	1
3	Tied to charger for conversations lasting more than 45 minutes.MAJOR PROBLEMS!!	0
4	The mic is great.	1
5	I have to jiggle the plug to get it to line up right to get decent volume.	0
6	If you have several dozen or several hundred contacts, then imagine the fun of sending each of them one by one.	0
7	If you are Razr owner...you must have this!	1
8	Needless to say, I wasted my money.	0
9	What a waste of money and time!.	0

```
In [11]: # Verify the distribution of the data
sns.countplot(data=df, x='sentiment')
```

```
Out[11]: <AxesSubplot:xlabel='sentiment', ylabel='count'>
```



There are slightly more positive reviews than negative reviews.

Data Cleaning

```
In [12]: # Check that the data type of the sentiment column is an integer
df['sentiment'].dtype
```

```
Out[12]: dtype('int64')
```

```
In [13]: # Drop null values
df = df.dropna()
```

```
In [14]: # Convert any non string values into string values
for i in range(0, len(df)-1):
```

```
if type(df.iloc[i] ['review'])!=str:
    df.iloc[i] ['review'] = str(df.iloc[i] ['review'])
```

```
In [15]: # Remove unusual characters: uppercase letters, punctuation marks, emoticons, and non-E
reviews = df['review']
list_of_characters = []
for r in reviews:
    for char in r:
        if char not in list_of_characters:
            list_of_characters.append(char)
print(list_of_characters)
```

```
['S', 'o', ' ', 't', 'h', 'e', 'r', 'i', 's', 'n', 'w', 'a', 'y', 'f', 'm', 'p', 'l',
'u', 'g', 'U', 'I', 'b', 'c', 'v', '.', 'G', 'd', ' ', 'E', 'x', 'j', 'T', '4', '5',
'M', 'A', 'J', 'O', 'R', 'P', 'B', 'L', '!', 'z', 'N', 'W', 'q', 'H', '+', 'V', '"',
'Y', 'D', 'F', 'k', '"', 'K', 'C', '/', '7', '3', '6', '8', '0', '2', '?', 'Z', '-',
'1', ':', ')', '(', 'Q', '&', '$', '*', ';', 'X', '%', '9', '#', '[', '']
```

```
In [16]: # Tokenize sentences into a list of tokens and remove any stop words
```

```
In [17]: review_list = []
stop_words = stopwords.words('english')
for review in df.review:
    review=re.sub("[^a-zA-Z0-9]", " ",review) # Use Regular Expression to remove non-alf
    review = review.lower() # Convert all characters to lower case
    review=nlTK.word_tokenize(review) # Tokenize the description
    lemma = nltk.WordNetLemmatizer() # Perform Lemmatization
    review = [lemma.lemmatize(word) for word in review]
    review = [word for word in review if not word in stop_words] # Remove stop words
    review = " ".join(review) # Join the description words
    review_list.append(review) # Add the clean description to the description_list
print(review_list)
```

```
['way plug u unless go converter', 'good case excellent value', 'great jawbone', 'tied c
harger conversation lasting 45 minute major problem', 'mic great', 'jiggle plug get line
right get decent volume', 'several dozen several hundred contact imagine fun sending one
one', 'razr owner must', 'needle say wasted money', 'waste money time', 'sound quality g
reat', 'wa impressed going original battery extended battery', 'two seperated mere 5 ft
started notice excessive static garbled sound headset', 'good quality though', 'design o
dd ear clip comfortable', 'highly recommend one ha blue tooth phone', 'advise everyone f
ooled', 'far good', 'work great', 'click place way make wonder long mechanism would las
t', 'went motorola website followed direction could get pair', 'bought use kindle fire a
bsolutely loved', 'commercial misleading', 'yet run new battery two bar three day withou
t charging', 'bought mother problem battery', 'great pocket pc phone combination', 'owne
d phone 7 month say best mobile phone', 'think instruction provided helpful', 'people co
uldnt hear talk pull earphone talk phone', 'hold charge', 'simple little phone use break
age unacceptable', 'product ideal people like whose ear sensitive', 'unusable moving car
freeway speed', 'two year left contract hate phone', 'car charger well ac charger includ
ed make sure never run juice highy recommended', 'need least 3 min get phone book time f
irst turn phone battery life short', 'ha kept well', 'poor talk time performance', 'case
great work fine 680', 'worthless product', 'ha great camera thats 2mp pic nice clear gre
at picture quality', 'wa impressed product', 'nice headset priced right', 'hear garbage
audio', 'excellent bluetooth headset', 'ha feature want', 'right mind gon na buy batter
y', 'arguing verizon regarding dropped call returned phone two day', 'case seems well ma
de', 'disappointed battery', 'loud enough turn like', 'good protection doe make phone bu
lky', 'usable keyboard actually turn pda real world useful machine instead neat gadget',
'phone pretty sturdy never large problem', 'love thing', 'everything fine reasonable pri
ce e', 'disappointed', 'even dropped phone stream wa submerged 15 second still work grea
t', 'happy 510 complaint one regarding sound quality end', 'button bad', 'essentially fo
```

rget microsoft tech support', 'really recommend faceplate since look nice elegant cool',
 'headphone great find think perhaps best purchase made last several year seriously', 'bu
 y different phone', 'hold phone particular angle party hear clearly', 'one big drawback
 mp3 player button phone front cover let pause skip song lock second', 'week later activa
 ted suddenly died', 'feel comfortable headset wear glass get way sometimes', 'great ipo
 d', 'device great several situation 1', 'mic work', 'even bmw 3 series fairly quiet trou
 ble hearing person saying', 'great choice', 'nice docking station home work', 'beautiful
 phone', 'love product', 'buy d807 wrongly advertised d807', 'great little item', 'love p
 hone handy ha lot feature', 'purchased within 2 day wa longer working', 'wear everyday h
 old well', 'good bargain', 'well packaged arrived time work intended', 'battery run quic
 kly', 'item worked great broke 6 month use', 'headset easy use everyone love', 'phone wo
 rk great', 'construction headset poor', 'product wa excellent work better verizon one bo
 y wa cheaper', 'phone load super', 'product cost much one doe expect work far better gre
 ater ease thing doe', 'made sturdy', 'worked great', 'ear bud play music one ear', 'dont
 waste money', 'order', 'plan use car forget', 'found product waaay big', 'best found far
 tried 2 bluetooths one ha best quality listener well ease using', 'disappointed decisio
 n', 'integrated seamlessly motorola razr phone', 'buyer beware could flush money right t
 oilet', 'definitely bargain', 'price wa good free shipping wa good purchase', 'received
 supposedly new motorola 375 apparently wa match phone charger', 'pro good camera nice pi
 cture also ha cool style like black white', 'huge design flaw unless using correctly thi
 nk case', 'love 350 headset jabra350 bluetooth headset great reception good ear piece co
 mfortable fit', 'camera although rated impressive 1 3 megapixels render image fall well
 expectation relatively high resolution', 'purchased car charger doe work', 'sound qualit
 y end excellent use headset call wife ask wife use headset call 2', 'phone slim light di
 splay beautiful', 'little geeky think thats sex toast rock ooze sex right battery embedd
 ed sleek stylish leather case', 'great product fast shipping', 'absolutely great', 'keyb
 oard nice compromise full qwerty basic cell phone number keypad', 'got phone around end
 may completely unhappy', 'comfortable ear', 'ha winner u', 'setup simpler', 'tried sever
 al different earpiece cell phone jabra one first one found fit ear comfortably', 'could
 get strong enough signal', 'iam pleased purchase', 'bad quality', 'get job done', 'great
 audio', 'basically service wa bad', 'work cell phone plug set charger', 'bluetooth week
 really like', 'bt headset wa disapointment', 'small even realize getting used', 'accompa
 nied software almost brilliant', 'headset work great wa packaged nicely avoid damage', 'd
 efinintly recommend', 'buyer careful', 'majority logitech earbud headset failed', 'good s
 tuff price', 'headset work peachy keen', 'house wa getting dropped coverage upstairs cov
 erage basement', 'voice recognition tremendous', 'waste', 'several different 2 minute ca
 ll experienced 5 drop wa area 5 bar', 'wa happy item', 'disappointed', 'phone take forev
 er charge like 2 5 hour literally', 'got phone reccomendation relative glad', 'waste tim
 e', 'reception poor', 'item stated included description included', '3 month screen went
 black sudden', 'wa hoping', 'audio quality poor poor', 'absolutely problem headset linki
 ng 8530 blackberry curve', 'know sound funny seemed like sketchy technology work well we
 ll one work great', 'bought 5 wired headphone sound better', 'phone fast sending kind me
 ssage web browsing significantly faster previous phone used', 'nice quality build unlike
 cheap', 'item fantastic work perfectly', 'love color', 'w810i superb', 'whine internet g
 o use thing le like', 'great charger', 'essentially communication tool doe communicate',
 'work like charm work one got phone', 'still maintain monkey make headphone obviously sh
 are enough dna copy design human', 'great value', 'work bough 17c working', 'look sharp
 screen nice clear great graphic', 'last long', 'one thing hate mode set button side', 'e
 xcellent service', 'thank wasting money', 'audio quality poor', 'igo charger tip really
 great', 'file browser offer option one need handsfree great', 'superb phone great networ
 k', 'connected wife bluetooth motorola hs850 phone worked like charm whether phone wa po
 cket case', 'put latest v1 15g like slow crawl lock every', 'recognizes phone storage de
 vice', 'like loud buzzing override conversation phone', 'buy product', 'bluetoooth heads
 et time still comfortable way fit ear', 'sound quality good functionality awesome', 'pho
 ne thorn side really abhor', 'recently problem could stay connected 10 minute disconnect
 ed', 'incredible', 'design flaw', 'great phone', 'waste 13 buck', 'check voice mail nigh
 t keypad backlight turn second first message lost', 'replacement died week', 'ring tone
 overall good phone buy', 'lately extremely nice helpful phone', 'problem wit hit droppin
 g signal', 'impressed', 'light weight hardly notice', 'love thin', 'bad choice', 'pleath
 er case fit', 'useless phone simply deaf', 'color even prettier thought would graphic in
 credibly sharp', 'think wa good investment', 'make strange ticking noise end call', 'bes
 t electronics available fm transmitter', 'battery last hour', 'great headset impressed h
 500', '1 2 mega pixel camera part phone reasonably good7', 'nearly 2 year ha worked grea
 t', 'good transmit audio', 'sending back', 'bother contacting company dollar product lea
 rned lesson bought form online anyway', 'hate earbugs avoid phone mean', 'great price al

so', 'range decent able roam around house phone living room reception sound quality issue', 'fast service', 'would highly recommend', 'would recommend', 'also like button felt like would crack use', 'worst phone ever', 'still infatuated phone', 'freeze frequently 4', 'kind embarrassing use look mostly embarrassing child like company', 'simple lightweight great fit', 'expected better consumer experience motorola', 'horrible tick sound background call never experienced', 'certainly comfortable functionality decent', 'usually like headband one lightweight mess hair', 'bit disappointed', 'phone year tell great', 'ordered sony ericsson w810i think worked thats first used', 'battery completely useless', 'big waste time', 'waste one', 'great sound service', 'really pleased product far', 'one favorite purchase ever', 'best bluetooth market', 'authentic leather nice shine comfort recommend case', 'was excited get headset thought was really cute', 'make mistake', 'big disappointment calendar sync', 'phone was unusable was new', 'worst customer service ever', 'additional ear gel provided instruction whatsoever', 'defeat purpose bluetooth headset', 'worth every penny', 'excellent wallet type phone case', 'nice headphone price work great', 'internet excruciatingly slow', 'comfortable wear well probably important aspect using case', 'fit like glove strong secure durable', 'gosh best phone ever', 'attractive appears good quality', 'love feature form factor', 'product strong rubber petroleum smell was unbearable caused return', 'poor quality service', 'cable look thin flimsy scary', 'great earpiece', 'excellent hand free tool', 'great using home stereo', 'good product incredible value', 'month worked flawlessly far', 'work great', 'absolutely junk', 'one work either', 'battery life real good', 'returned 8 hour later', 'great item', 'waste money', 'love headset', 'drain player may also potentially fry', 'worked perfectly', 'unreliable giving', 'gave 5 star sound quality', 'avoid one', 'reversible plug work great', 'wasted little money earpiece', 'good phone', 'best rotating feature helpful', 'comfort whole family', 'good product good seller', 'love bluetooth', 'first wear well', 'impressed headset plantronics', 'got was small adorable', 'great phone', 'recommend buying product', 'piece junk', 'poorly constructed hinge', 'item great installed let charged overnight has working good every since problem', 'great phone', 'buy', 'excellent phone', 'purchase phone', 'voice recognition thru handset excellent', 'hand best phone ever', 'work great cat attacked phone scratched protective strip instead destroying screen', 'one work was priced right', 'customer service was terrible', 'battery life still long enough motorola razor v3i', 'sending back', 'know made wise decision', 'someone should've invented sooner', 'good price', 'everything worked first try device was certainly engineered clever way construction feel good', 'received quickly work great', 'has working great', 'many people complained found headset microphone was weak', '5 year old nokia 2160 tracfone hold charge lot better', 'care consumer one bit', 'good battery got really fast', 'instruction manual lacking', 'also slim enough fit alarm clock docking station without removing case', 'phone hold charge like was new', 'battery unreliable well service use antenna', 'uncomfortable sound quality quite poor compared phone razr previous wired headset plugged in', 'gotten lot compliment', 'definitely must state don't allow cell phone usage driving', 'want money back', 'love device', 'also happy price', 'ordered product first was unhappy immediately', 'engage still lacking earbuds', 'was able voice dialing car problem', '2 3 bar cell phone home can't hear anything', 'good item low price', 'however ringtones best neither game', 'plug work well', 'amazon suck', 'rip charge shipping', 'comfortable', 'battery working well replacement original came phone 2 year ago', 'reception excellent', 'fit comfortably either ear sound clear loud charge last couple days', 'clear quality sound mess sound ipod since sound button headset', '1 long lasting battery recharge frequently flip phone 2', 'obviously problem adhesive', 'disappointing', 'was inexpensive piece would still expected better quality', 'practically useless add kind boost reception bought', 'excellent sound quality', 'was happy', 'piece junk broke phone 2 days', 'new battery work great phone', 'worst customer service', 'dropped phone time say even concrete phone still great knock wood', 'reception terrible full static', 'good audio quality', 'little device has transformed organizational capability made life whole lot easier', 'was sitting vehicle cradle belt headset lost signal', 'jerk phone', 'work like charm work advertised', 'comfortable nice range good battery life', 'network coverage los angeles horrible', 'excellent starter wireless headset', 'loudspeaker option great bumper light appealing', 'thing think could improve sound leak headset', 'first reception suck never 2 bar ever', 'phone get extremely hot', 'microphone also work well according people called applies everything', 'nice design quality', 'also phone dropped case going save specially dropped face', 'light easy use has clear reception transmission', 'bad fit way big', 'definitely was good still', 'awesome device', 'waste money', 'ordered motorola data cable got well finished working product', 'love phone', 'was looking headset long time got happier', 'driving along headset start ringing reason', 'display excellent camera good year', 'battery life also great', 'worst phone ever month', 'good item worked started problem auto reverse tape player', 'embarrassing also ear hurt try push ear plug ear', 'protects phone side

e', 'average phone bad battery life operates weak network', 'clear skype call long battery life long range', 'soyo technology suck', 'great hand free device', 'even take self portrait outside exterior display cool', 'problem others mentioned', 'trying many many handsfree gadget one finally work well', 'magical help', 'worst piece crap ever along verizon customer service', 'poor sound quality', 'best phone market', 'worked well', 'company shipped product promptly work well', 'exactly wanted', 'picture resolution far comparably priced phone offering today', 'great deal', 'excellent product satisfied purchase', 'highly recommend encourage people give try', 'better expect', 'using earpiece left right thing stay ear', 'great product price', 'sound good reasonably priced effective simple', 'still waiting sure item would work well ever receive', 'disappointment', 'prompt service', 'excellent product price', 'pleased headset', 'waste money time', 'especially love long battery life', 'stupid keep buying new charger car charger cradle headphone car kit every time new phone come', 'work great', 'plantronics bluetooth excellent buy', 'came verizon cingular ha nicer cell phone first thing noticed wa really bad service', 'looking new earpiece', 'highly recommend device everyone', 'jawbone era awesome', 'received white colored battery go dead couple hours the original used last week still last longer thereplacement', 'year battery went completely dead headset', 'used several phone two years one best', 'awesome new look fall 2000', 'first phone ha cheaply made', 'att clear sound distorted yell talk', 'plastic break really easy clip', 'price good', 'make mistake', 'oh forgot also mention weird color effect ha phone', 'using iriver spinn case fit fine', 'also area unit broke fond magnetic strap', 'overall psyched phone ha appointment contact get great reception', 'every thing phone work perfectly like', 'another note phone appearance really look rather bland especially black model', 'really easy', 'sanyo ha survived dozen drop blacktop without ill effect', 'fantastic earphone', 'find cell phone right away enter car', 'low quality', 'highly recommend modest priced cellular phone', 'doe get better reception clarity phone', 'warning stay away', 'wish could return unit get back money', 'awsome device work great', 'used talk 1 1 1 2 hour battery would literally drained dying', 'find bad unit come one earpad would recommend anyone', 'displeased', 'might gotten defect would risk buying built quality alone', 'made case small difficult install', 'restored phone like new performance', 'would recommend purchasing jabra jx 10 series 2 work flawlessly moto q go figure', 'searched internet found one best value', 'doe fit', 'happy product', 'screen size big key pad lit well enough camera quality excellent camera phone', 'really nothing bad say headset', 'glad found product amazon hard find high priced', 'portable work', 'colleague get great reception little expensive performance great', 'even fully charged went bed turned blue tooth wi fi noticed 20 left morning', 'also problem phone reading memory card always turn', 'good wearing hat sunglasses', 'timely shipment quality product', 'nice solid keyboard', 'well made fit surefire gx2 perfectly', 'horrible phone', 'bt50 battery junk', 'would recommend others', 'cant get software work computer', 'disappointed', 'poor voice clarity', 'battery work great', 'buyer remorse one', 'disappointed accessory one', 'find inexcusable probably returning phone perhaps changing carrier', 'look great strong', 'tmobile service 2 3 year pretty happy', 'update procedure difficult cumbersome', 'disappointment hate anything go ear', 'great software motorolas', 'delivery wa time', 'easy use', 'uncomfortable ear use lg vx9900 env', 'volume switch rocketed unit destination unknown', 'hold charge long time reasonably comfortable long wearing condition quality sound tremendous', 'like return', 'keyboard really worthwhile usefulness sturdy enough expect problem', 'verizon bill however difficult understand even though pricing plan simple', 'charging overnight battery work great', 'regret', 'love phone', 'disappointment', 'user friendly', 'unfortunately ability actually know receiving call rather important feature phone pitiful respect', 'exchanged phone even problem', 'result phone either stuck max volume mute', 'pda rock', 'super charged use small hybrid palmtop camera cellphone excels role', 'would definitely recommend jabra bt250v looking comfort clarity great price', 'verizon 2 year ago really liked service', 'pretty piece junk', 'plug wa wrong size', 'work described', 'save money item 11 months', 'bad purchase', 'well im satisfied', 'bose noise cancelling amazing important nyc commuter', 'battery excellent bargain', 'defective crap', 'would given star wa able', 'wear glass fit fine', 'sound quality device unacceptable unless really quiet area almost hear nothing', 'would recommend item anyone', 'holster arrived match photo ad', 'great problem', 'earlier review noted plug charger nothing happens', 'frog phone charm adorable eye catching', 'never know pushed hard enough right number time function want', 'sound clear people talk amazed quality', 'got aluminum case new palm vx worked really well ha protected handheld perfectly far', 'good product well made', 'mobile phone tool must', 'impressed job motorola sturdiness phone', 'received order well within shipping timeframe everything wa good working order overall excited source', 'phone sturdy waterproof', 'impressed would recommend item anyone', 'complaint standard sound volume little low even turned 5 5', 'thanks amazon thing need good price', 'ended sliding edge pant back pocket

instead', 'go store bought new nokia phone working great', 'work great', 'really ugly', 'horrible switch 3 time', 'noise shield incredible', 'real improvement', 'razr battery good buy', 'keep good work amazon', 'refuse refund replace', 'unfortunately easy accidentally activate gentle touch button accidentally touch phone face listening', 'basically threw money window nothing', 'great case price', 'took wa one drop 6 inch kitchen counter wa cracked impressed laughing', 'bluetooth range good day ago left phone trunk got call carried conversation without hitch', 'small sleek impressive looking practical setup ample storage place', 'earbuds channel sound directly ear seem increase sound volume clarity', 'none new one ever quite worked properly', 'glad purchased', 'missed numerous call reason', 'perhaps phone defective people hear use', 'buy product', 'sucked stuff doesn't work phone', 'shifting bubbling peeling even scratch nothing happier new one droid', 'worst software ever used could give zero star would', 'trying make call exercise frustration', 'case flimsy piece plastic has front side protection whatsoever', 'jabra earset was happy sound quality especially outgoing better', 'love camera really pretty good quality', 'thank great service', 'total package', 'seller understanding patient would definitely buy', 'give wirefly 1 star contact cingular inform practice', 'aggravating', 'great product price', 'purchased family friend enjoy clarity ease use', 'virgin wireless rock doesn't cheap little phone', 'none work buy', 'muddy low quality sound casing around wire insert was a poorly super glued slider', 'please don't waste money', 'quite satisfied purchase', 'looking good quality motorola headset keep looking', 'yet plantronics continues use flawed charger design', 'disappointing result', 'cumbersome design', 'great was new packaged nice work good problem came later time expected', 'great product', 'piece crap', 'experience was terrible was fourth bluetooth headset was much comfortable last jabra hated', 'fix problem', 'worst ever', 'happy', 'trouble accessing internet downloading ringtones performing function', 'barely ever hear constantly saying', 'battery life highly unacceptable', 'company joke', 'doesn't everything description said would', 'due happening every call was forced stop using headset', 'work great car charger especially plug two adapter time', 'verizon tech support walked procedure none worked ended hard set wiping data', 'voice quality signal strength dropped call', 'play louder speaker size price low would think quality lacking however', 'better expected', 'feel poorly constructed menu difficult navigate but ton recessed difficult push', 'battery holding well', 'also difficult put recommend avoiding product', 'used two day always easy hear', 'got car charger even week charger was broken went plug started smoking', 'sprint terrible customer service', 'linked phone without effort', 'cool', 'easy use', 'better new', 'possessed get junk idea', 'piece trash', 'pretty good product', 'great phone', 'price amazon excellent product would highly recommend', 'research development division obviously know', 'real killer volume course breaking', 'phone working fine', 'happy product', 'poor construction', 'infuriating', 'overall would recommend phone new walkman', 'sprint charge service', 'cheap feel look cheap', 'fit ear well comfortable', 'work europe asia', 'clipping belt definitely make feel like 50 cent coming', 'like design look jabra behind ear headset 5020 pretty comfortable wear 24 hour day without pain ear', 'was amazed quick arrival two original lg cell phone battery fraction price', 'worst phone samsung crap crappy samsung e715 worst ever seen', 'charger worked week completely stopped charging phone', 'poor product', 'however needed better instruction', 'need two hand operate screen software interface decade old compete new software design', 'worked well paired treo 700w skype headset pc using usb bluetooth transceiver', 'steer clear product go genuine palm replacement pen come three pack', 'buy buyit suck', 'great product sure beat using finger', 'plus seriously believe worth steep price point', 'case tried normally fall apart month one seems long haul', 'disappointing performance', 'make mistake please buy phone', 'phone might well worst ever brand', 'case was late extra originally put later discarded scratched ear', 'recommend if anyone different brand cell phone mp3 player family', 'bother go store', 'posted detailed comment grey black phone fire red great color', 'problem item would order needed', 'get pay guess', 'software great job could also make ringtones from existing cd without internet connection', 'im surprised good quality car charger much review', 'fabulous', 'however bluetooth headset currently good real time game like first person shooter since audio delay mess', 'used bitpim free program find internet transfer data phone price cable was excellent', 'disappointing accessory good manufacturer', 'performed awful muffled tinny incoming sound severe echo end call', 'wind resistant', 'obviously terrible customer service get pay', 'last forever overly expensive replace easy operate sound much better others tried', 'contacted company told although unit was still warranty produce receipt was luck', 'linksys way exchange bad phone refurb unit something', 'phone sturdy nokia bar phone', 'great phone', 'placed treo case was snug was lot extra room side', 'pleased', 'battery terrible', 'found product easy set use', 'earpiece large heavy keep falling ear', 'utter crap sound quality terrible', 'charger arrived within promised timeframe work', 'ear loop needed tiny sound great', 'finally three four time spring latch broke could use longer visor

r', 'handsfree part work fine car try download address book treo reboots overall still r
ate device high', 'talk useless customer service', 'felt light tinny', 'doe fit palm tun
gsten e2 broke first time tried plug', 'happy purchase good quality low price', 'love lo
ok feel samsung flipphones', 'well designed sharp blue nice color', 'doe everything', 'c
ould recommend', 'nice sound', 'look good picture case wa huge disappointment', 'much le
jawbone wa going replace', 'setup went smoothly', 'extremely slow take forever anythin
g', 'must study anyone interested worst sin industrial design', 'first thing happened wa
tracking wa', 'well made easy access phone ha handy detachable belt clip', 'continue pai
ring periodically since somehow kept dropping', 'upload ringtones third party', 'bluetoo
th doe work phone lock screen flash make call randomly pocket locked', 'truly awful', 'w
ork fine motorola 325 cellphone lot better worn battery', 'volume ringer real good choic
e loud', 'none tone acceptable', 'camera color balance awful', 'product clearly ready pr
ime time care cool look tell call coming worthless', 'love ringtones upbeat', 'chinese f
orgery abound', 'easy pair samsung cell', 'instruction explain microphone jack could use
d', 'would advise purchase item never worked well', 'bad pay month service', 'work wel
l', 'nokia ca 42 usb cable work phone', 'clear crisp sound', 'use smallest earpiece prov
ided stay pretty well', 'biggest complaint battery drain superfast', 'design might ergon
omic theory could stand ear', 'buy want use holster', 'phone also take great picture eve
n video clip', 'make look cool', 'ear occupied background distracting', 'comfortable fit
need headset comfortable least hour time entire day', 'also phone seem accept anything e
xcept cbr mp3s preferably ripped window medium player', 'beat price', 'come strong light
use light camera shot even flash signal seriously', 'lightweight work well', 'beware',
'love cable allows connect mini usb device pc', 'anyone near hear part conversation', 'c
ool case open allowing screen get scratched', 'bought phone replacement startac regrette
d since', 'appears actually outperform original battery china came v325i', 'store anythi
ng phone number sim', 'last le 30 minute actually try use phone wife ha phone problem',
'poor quality', 'jabra ear gel r best', 'worth', 'phone crashed completely get replace
d', 'quit working used 18 month purchased another one best headset ever owned', 'unfortu
nately recharge iphone 4 despite connecting multiple power source imac external battery
wall outlet etc', 'wa great phone', 'simply want small flip phone look elsewhere extra b
ell whistle mediocre', 'good value work fine power via usb car wall outlet', 'phone slid
e around car grip prevents phone slipping hand', 'simply best bluetooth headset sound qu
ality', 'think plan one sure last moto phone', 'span hour two people exclaim whoa new ph
one tv', 'always used corded headset freedom wireless helpful', 'case ha passed one year
mark show sign wear 100 functional', 'leather nice soft fit tight cut face good shape',
'ha everything need ask', 'receiving using product 2 day broke', 'product great make wor
king lot easier go copier waiting hold something', 'none three size sent headset would s
tay ear', 'mobile ha best customer service anywhere', 'arrived quickly much le expensive
others sold', 'tried 2 unit failed within 2 month pro', 'provides good protection look c
lassy', 'krussel always good', 'tracfonewebsite user friendly make easy toactivate etc',
'design good 4', 'great reception time', 'would highly recommend product', 'use unit sun
glass good texas', 'dit work time nokia 5320', 'bought mainly charger broke soon purchas
ing', 'like loud buzzing override conversation phone', 'fantastic buy get whatever next
phone', 'stay away store careful', 'blueant supertooth hand free phone speaker awesome',
'doe fit', 'item arrived quickly work great metro pc samsung sch r450 slider phone sony
premium sound ear plug', 'light ha plenty battery capacity confortable wear somewhat ext
ended period time', 'disappointing', 'blue ant easy use', 'cheap hey work wa pleasantly
suprised given low cost item', 'camera phone may used dustpan indoors rather using dispo
sable', 'day later puff smoke came phone use', 'convenient simple use get job done make
car ride much smoother', 'nano stated son wa dissapointed', 'sound quality excellent wel
l', 'best headset used', 'would recommend', 'around year carry highest quality anti gla
re screen protector found date', 'wanted', 'particular model would work motorola q smart
phone', 'wont work right atleast', 'addition feel amp look phone lightweight cheap plast
ic', 'warranty problem reoccure bottom line put money somewhere else cingular support',
'really wanted plantronics 510 right one ha many issue good', 'great product', 'crisp cl
ear', 'first charge kept going dead 12 minute', 'cheap plastic creak like old wooden flo
or', 'get absolutely horrible reception apartment phone problem', 'reception ha generall
y good', 'excellent sound battery life inconspicuous boot', 'great bluetooth', 'highly r
ecommend case', 'worthwhile', 'great deal', 'phone got year old slowly breaking despite
much care part', 'sorry made purchase', 'bought hoping could make bluetooth headset fit
better thing made impossible wear', 'treo mobile refused replace forced buy another phon
e without kind upgrade discount', 'wa horrible', 'think would hold securly belt', 'allow
s possibility double booking date time first', 'work 2 week break', 'simple use like',
'great tool entertainment communication data management oh sure use activesync 4 2 optim
al data synchronization result', 'use case smell disgusting', 'wanted', '2 case would or

der another', 'great phone', 'bought battery coupon amazon happy purchase', 'internet access wa fine rare instance worked', 'perfect ps3', 'five star plus plus', 'good quality bargain bought bought cheapy big lot sounded awful people end hear', '50 drain', 'tried talking real loud shouting telephone get old wa still told great', 'good work fine', 'price wa great deal', 'great price', 'wind completely useless', 'yes shiny front side love', 'doe work', 'first thought wa grtting good deal 7 44 plugged phone v3c razr', 'reception headset excellent', '2 thumb seller', 'verizon lg phone work well together good reception range exceeds 20 foot line sight', 'improper description return', '1 work 2 comfortable', 'battery charge life quite long', 'jabra eargels fit ear well', 'battery ha life', 'checked everywhere feature really disappointing', 'product exactly described', 'get signal verizon phone', 'job', 'would recommend', 'plan ordering', 'said phone camera nice many cool effect play video decent well', 'satisfied cell phone day one', 'keep getting better better third one numerous palm', 'awkward use unreliable', 'shipping time wa also quick', 'good hoped', 'good case', 'work', 'really good product', 'unfortunately work', 'three broke within two month use', 'dont buy', 'father ha v265 battery dying', 'however ear pad come easily one week lost one', 'call drop phone come screen go black worst stop ringing intermittently', 'horrible horrible protector', 'like nokia either', 'reaching bottom row uncomfortable send end key expect 3', 'bought nokia cell phone week ago wa a nightmare', 'would take long describe bad customer service experience ha amazon', 'disappointing thing wa wa speakerphone', 'terrible car accept cassette', 'product useless since doe enough charging current charge 2 cellphone wa planning use', 'say', 'stay away q', 'used dirty', 'unreliable service mobile', 'good enough price', 'bought two neither charge', 'look nice', 'excellent tool especially paired phone auto answer', 'read review problem', 'product cover important light sensor ear outlet', 'comfortable hand', 'doe work', 'poor reliability', 'work', 'take picture battery drop bar start beeping letting know dieing', 'terrible product waste money', 'last time buying', 'wa waste money', 'far ha worked like charm', 'week far good', 'case best part download picture laptop using ir even send picture laptop phone', 'would recommend item', 'choice headset year great sound good volume good noise cancellation', 'nice case feel good hand', 'item received wa counterfeit', 'really good far probably best bt headset ever', 'work fine', 'received headset good time wa happy', 'wa pleased see could replace well travled swivel holster black berry flip', 'sister ha one also love', 'good quality', 'far worst purchase made amazon', 'excellent dual purpose headset', 'excellent', 'doe work listening music cingular 8125', 'seems completely secure holding belt keeping iphone inside', 'bottom line another worthless cheap gimmick sprint', 'trouble volume', 'good thing wa fit comfortably small ear', 'warning buy', 'pairing iphone could happier far', 'open battery connection broken device turned', 'want clip going top ear causing discomfort', 'best headset ever', 'trust website expect helpful support', 'wa loud glad say plantronics 510 maintains flawless connection cell static normal use', 'speaker low quality making ring tone sound cheap', 'fit comfortably came three size earbud tip', 'excellent product price', 'bad reception', 'product high quality chinese crap', 'wa quite comfortable ear', 'great phone', 'buy product fails', 'everything product wrong first', 'pairing two device wa easy barely took couple minute started making call voice dialing feature', 'wa utterly confusing first caused lose couple important contact', 'terrible phone holder', 'cutout button placed perfectly', 'love able use one headset land line cell', 'problem ear loop made weak material break easily', 'nearly good looking amazon picture make look', 'despite flaw case exceptional quality well worth additional cost owning official oem product', 'stupid', 'heard driving car usually even put loudest setting', 'seller shipped quickly much cheaper competitor', 'good show samsung', 'bought 10 saved alot money', 'always cut make beep beep sound say signal failed', 'figure', 'phone sounded ok great ok wife phone wa almost totally unintelligible understand word said', 'take battery phone put back together rest art', 'managed bend leaf spring back place metal ha enough stress break next drop', 'leopard print wonderfully wild', 'work', 'saggy floppy piece junk', 'much better hard plastic case', 'also new problem', 'happy far', 'cool phone', 'work good', 'motorola finally got voice quality bluetooth headset right', 'loos connection phone even within couple foot two', 'great would really recommend', 'make easier keep bluetooth wearing', 'said pretty useless', 'never got', 'pair headphone worst ever sound wise', 'drain weak snap', 'doe charge cingular att 8525 phone', 'get extra minute carry call get cut', 'however year flip top started get loose wobbly eventually screen went black receive place call', 'kept catching cable seat pull phone turn', 'headset fulfills requirement happy purchase', 'logitech bluetooth headset 10', 'like fact rest lightly ear rather inside', 'lot website rating good phone', 'tried cable computer ipod work fine', 'mic joke volume quite low', 'returning', 'doe maintain connection computer lap', 'easy turn car volume control quite accessible', 'mine year christmas bought rest family', 'otherwise easy install use clear sound', 'nice leather', 'purchase different case', 'joy use', 'much disappointed companion

y', 'satisfised', 'charge', 'phone get 2005 bought s710a say wow', 'button small', 'reading spec alone make say wow', 'love great armband', 'really like product motorola allot clearer ear piece mic', 'however keypad tinny sometimes reach wrong button', 'exchanged sony ericson z500a pretty happy decision', 'wa impressed price case', 'also make easier hold', 'use product motor control center lot high voltage humming equipment work great', 'pretty easy', 'make call certain place', 'however girl wa complain time phone wake like normal phone doe', 'beautiful styling though', 'company charge restocking fee still given refund back', 'trying return refund', 'plugged find darn thing worked', 'excellent product', 'earbud piece break easily', 'lousy product', 'phone try hard everything fails ability phone', 'best charger seen market yet', 'sweetest phone', 'oh charger seems work fine', 'fit securely ear hook doe even need used sound better directed ear canal', 'enough volume', 'echo problem unsatisfactory', 'could take 2 video time quality wa poor', 'waste money', 'going first negatively review product', 'adapter doe provide enough charging current', 'wa much hype phone assumed wa best mistake', 'also take picture case lense covered', 'phone fall easily', 'work people hear talk', 'text messaging feature really tricky use', 'really disappointed charger work', 'painful ear', 'lasted one day blew', 'disappointed', 'kind flop around', 'screen doe get smudged easily touch ear face', 'piece junk lose call phone', 'item doe match picture', 'thing disappoint infra red port irda', 'answer call unit never worked']

```
In [18]: # Identify vocabulary size - bag of words
tokenizer = Tokenizer()
tokenizer.fit_on_texts(df['review'])
vocabulary_size = len(tokenizer.word_index) + 1
print("Vocabulary Size: ", vocabulary_size)
```

Vocabulary Size: 1879

There are 1879 unique words in the reviews. This is our vocabulary size.

```
In [19]: # proposed word embedding length - Determine the word embedding and explore embedded words
# Source: https://machinelearningmastery.com/use-word-embedding-layers-deep-learning-keras/
word_embedding_length = int(round(numpy.sqrt(vocabulary_size)))
print("Proposed word embedding length:", word_embedding_length)
tf.keras.layers.Embedding(vocabulary_size, word_embedding_length)
```

Proposed word embedding length: 43

```
Out[19]: <keras.layers.core.embedding.Embedding at 0x245fa56f3d0>
```

The proposed word embedding length is determined by the distance of a word from the beginning of the vector. For example, if in a vector that has 300 spaces, the position of the word is 120 counting from the beginning of the vector, the embedding length of that word is 120 (Brownlee, 2021).

I have used the process of the 4th square root to determine the proposed length.

```
In [20]: # statistical justification for the chosen maximum sequence length
# Set the maximum sequence length based on the length of the longest sentence in the dataset
max_sequence_length = 0
sequence_lengths = []
for review in review_list:
    review_length = len(review)
    sequence_lengths.append(review_length)

# Calculate maximum sequence length
max_sequence_length = np.max(sequence_lengths)

# Calculate median sequence length
```

```

median_sequence_length = np.median(sequence_lengths)

# Calculate minimum sequence length
min_sequence_length = np.min(sequence_lengths)

print(f"Maximum sequence length: {max_sequence_length}")
print(f"Median sequence length: {median_sequence_length}")
print(f"Minimum sequence length: {min_sequence_length}")

```

```

Maximum sequence length: 117
Median sequence length: 29.0
Minimum sequence length: 3

```

Now we have the information for padding sequences. We know that our minimum sequence will be padded with 114 tokens in order to meet the maximum length sequence in order to fit all the sequences into a matrix.

B2. Describe the goals of the tokenization process, including any code generated and packages that are used to normalize text during the tokenization process.

Rubric: The submission completely describes the goals of the tokenization process, including any code generated and packages that are used to normalize text during the tokenization process. The description is accurate.

Resource: https://tensorflow.org/api_docs/python/tf/keras/preprocessing/text/tokenizer

The TensorFlow tokenizer class was used to vectorize the text corpus, by turning each text into a vector of integers. The vocabulary size was then created from the tokenized vector by counting the length of the indexes on the vector. The vocabulary size was then used to find the word embedding length. The word embedding length is important to building our NLP model as it is an input value for one of the parameters.

B3. Explain the padding process used to standardize the length of sequences, including the following in your explanation:

Rubric: The submission accurately explains the padding process used to standardize the length of sequences and includes an explanation of where the padding occurs. The padding process aligns with the chosen data, and a screenshot of a single padded sequence is provided.

- if the padding occurs before or after the text sequence
- a screenshot of a single padded sequence

In order for the NLP model to run as expected all sequences must be the same length. Padding of the sequences makes sure all the sequence lengths are the same size. I have chosen to pad and truncate the sequences after (post) the end of the sequence.

```

In [21]: sequences = tokenizer.texts_to_sequences(df['review'])
         padded_sequences = pad_sequences(sequences, maxlen=max_sequence_length, padding="post",

```

```

In [22]:

```

```
print(padded_sequences[1])
```

```
[ 21  48  54 272   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   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   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]
```

B4. Identify how many categories of sentiment will be used and an activation function for the final dense layer of the network.

Rubric: The submission clearly and accurately identifies how many categories of sentiment will be used and an appropriate fitting activation function for the final dense layer of the network.

Two categories of sentiment will be used. The activation function for the final Dense layer of the network is 'sigmoid'. Sigmoid is the best activation function to use on the final Dense layer of a binary sentiment problem. (<https://towardsdatascience.com/an-easy-tutorial-about-sentiment-analysis-with-deep-learning-and-keras-2bf52b9cba91>)

B5. Explain the steps used to prepare the data for analysis, including the size of the training, validation, and test set split.

Rubric: The submission explains the steps used to prepare the data for analysis and accurately includes the size of the training, validation, and test split. The steps relate to preparing for neural network models and NLP techniques.

1. Read the data from the dataset
2. Check on unusual or any abnormal characters
3. Clean the data
4. Explore the data
5. Create a 2-Dimensional array with reviews and sentiments by encoding
6. Split the data into train and test data sets with an 80/20 split
7. Apply the tokenizer to the training set with the fit_on_texts() method
8. Retrieve the word_index of the training set
9. Add post-padding to the sequences to get the maximum length needed
10. Fit the final model and convert both the training and test set into arrays using numpy

In [23]:

```
# Split the data set into training, test, and validation data sets
X = padded_sequences
y = numpy.array(df['sentiment'])

from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=.2, random_state=40)
X_train, X_validation, y_train, y_validation = train_test_split(X, y, test_size=.2, ran

print("Training Size:", X_train.shape)
print("Testing Size:", X_test.shape)
print("Validation Size:", X_validation.shape)
```


Training Size: (800, 117)
 Testing Size: (200, 117)
 Validation Size: (200, 117)

B6. Provide a copy of the prepared dataset.

Rubric: The submission provides a copy of a fully prepared dataset.

```
In [24]: pd.DataFrame(X_train).to_csv('./sentiment labelled sentences/X_training_df.csv')
pd.DataFrame(X_test).to_csv('./sentiment labelled sentences/X_testing_df.csv')
pd.DataFrame(X_validation).to_csv('./sentiment labelled sentences/X_validate_df.csv')
pd.DataFrame(y_train).to_csv('./sentiment labelled sentences/y_training_df.csv')
pd.DataFrame(y_test).to_csv('./sentiment labelled sentences/y_testing_df.csv')
pd.DataFrame(y_validation).to_csv('./sentiment labelled sentences/y_validate_df.csv')
```

Part III: Network Architecture

C. Describe the type of network used by doing the following:

C1. Provide the output of the model summary of the function from TensorFlow.

Rubric: The submission provides the complete output of the model summary of the function from TensorFlow. The output aligns with the type of network used.

```
In [25]: # Create an activation function filled with dense layers of the neural network
loss="binary_crossentropy"
optimizer="adam" #rmsprop

num_epochs=25

y_train.reshape((-1,1))
y_test.reshape((-1,1))
y_validation.reshape((-1,1))

# Define early stopping monitor
early_stopping_monitor=EarlyStopping(patience=2, verbose=1)

model = tf.keras.Sequential([
    tf.keras.layers.Embedding(vocabulary_size, word_embedding_length, input_length=max_
    tf.keras.layers.Bidirectional(tf.keras.layers.LSTM(20,dropout=0.6)), # Resource: ht
    #     tf.keras.layers.Dense(10, activation='relu'),
    #     tf.keras.layers.GlobalAveragePooling1D(),
    #     tf.keras.layers.Flatten(),
    tf.keras.layers.Dense(100, activation='relu'),
    #     tf.keras.layers.GlobalAveragePooling1D(),
    tf.keras.layers.Dense(50, activation='relu'),
    tf.keras.layers.Dense(2, activation='sigmoid'), # binary problems should have sigmo
    tf.keras.layers.Flatten()
])

model.compile(loss=loss, optimizer=optimizer, metrics=['acc'])
```

```
history = model.fit(X_train, y_train, batch_size=128, epochs=num_epochs, verbose=0, val
```

Epoch 14: early stopping

In [26]:

```
model.summary()
```

Model: "sequential"

Layer (type)	Output Shape	Param #
=====		
embedding_1 (Embedding)	(None, 117, 43)	80797
bidirectional (Bidirectional)	(None, 40)	10240
dense (Dense)	(None, 100)	4100
dense_1 (Dense)	(None, 50)	5050
dense_2 (Dense)	(None, 2)	102
flatten (Flatten)	(None, 2)	0
=====		
Total params: 100,289		
Trainable params: 100,289		
Non-trainable params: 0		

C2. Discuss the number of layers, the type of layers, and total number of parameters.

Rubric: The submission completely and accurately discusses the number of layers, the type of layers, and the total number of parameters in the network.

1st Layer: Core Type = Embedding

- Layer Type = Input
 - Total # Parameters = 80,797

2nd Layer: Core Type = Bidirectional

- Layer Type = Input
 - Total # Parameters = 10,240

3rd Layer: Core Type = Dense

- Layer Type = Hidden
 - Total # Parameters = 5050

4th Layer: Core Type = Dense

- Layer Type = Hidden
 - Total # Parameters = 102

5th Layer: Core Type = Flatten

- Layer Type = Output
 - Total # Parameters = 0

Total Number of Parameters

- 100,289 where 100,289 are trainable and 0 are non-trainable

C3. Justify the choice of hyperparameters, including the following elements:

Rubric: The submission logically justifies the choice of hyperparameters, including each of the 6 listed elements, and each element aligns with the network used.

- Activation Functions
 - Sigmoid was selected because it performs best for binary values such as those in the sentiment column of the data set. The formula for the Sigmoid activation function, $\text{sigmoid}(x) = 1 / (1 + \exp(-x))$. The second element is assumed to be zero. The sigmoid function always returns a value between 0 and 1. (Source: <https://keras.io/api/layers/activations/>)
- Number of Nodes Per Layer
 - The number of nodes that were selected for each Input Dense layers 3 and 4 was selected by trial and error.
 - The Output Dense layer was set to 2 nodes because the y values are sentimental. In other words 0 for a negative review and 1 for a positive review. This means there can only be two possible outcomes.
- Loss Function
 - The loss function was set to binary cross entropy because it is best for a binary outcome. "Loss: A scalar value that we attempt to minimize during our training of the model. The lower the loss, the closer our predictions are to the true labels. This is usually Mean Squared Error (MSE) as David Maust said above, or often in Keras, Categorical Cross Entropy." (<https://stackoverflow.com/questions/34673396/what-does-the-standard-keras-model-output-mean-what-is-epoch-and-loss-in-keras#:~:text=Loss%3A%20A%20scalar%20value%20that,in%20Keras%2C%20Categorical%20C>)
- Optimizer
 - The Adam optimizer was the selected optimizer because it is the most commonly used in a deep learning problem like we have built here. (<https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/>)
- Stopping Criteria
 - The stopping criteria is used in order to tell the network to stop early if a certain criteria is met. In this case I used the value of the accuracy score as my stopping criteria as that is what I was most interested in achieving.
- Evaluation Metric

- Accuracy was used as the evaluation metric to be used to assess the performance of the model as it was of the most interest.

Part IV: Model Evaluation

D. Evaluate the model training process and its relevant outcomes by doing the following:

D1. Discuss the impact of using stopping criteria instead of defining the number of epochs, including a screenshot showing the final training epoch.

Rubric: The submission accurately discusses the impact of using stopping criteria instead of defining the number of epochs. A screenshot showing the final training epoch is provided.

Resources: <https://stats.stackexchange.com/questions/231061/how-to-use-early-stopping-properly-for-training-deep-neural-network>, https://keras.io/api/callbacks/early_stopping/

The number or epochs is used to tell the model to continue to train within the range of 0 to the epoch set on the model. Epochs can be used to determine the best validation score. This can be done by a manual process or automatically by using the early stopping parameter called patience. Patience tells the model how many epochs the model should go without improving before training is stopped. A general rule is to set the patience to 2 or 3.

In [27]:

```
# List all data in history
# Sources: https://keras.io/api/callbacks/, https://machinelearningmastery.com/display-score = model.evaluate(X_test, y_test, verbose=1)
```

7/7 [=====] - 0s 29ms/step - loss: 0.4119 - acc: 0.8250

D2. Provide visualizations of the model's training process, including a line graph of the loss and chosen evaluation metric.

Rubric: The submission provides complete visualizations of the model's training process, including a line graph of the loss and the chosen evaluation metric. The visualizations are clearly labeled and align with the model's training process.

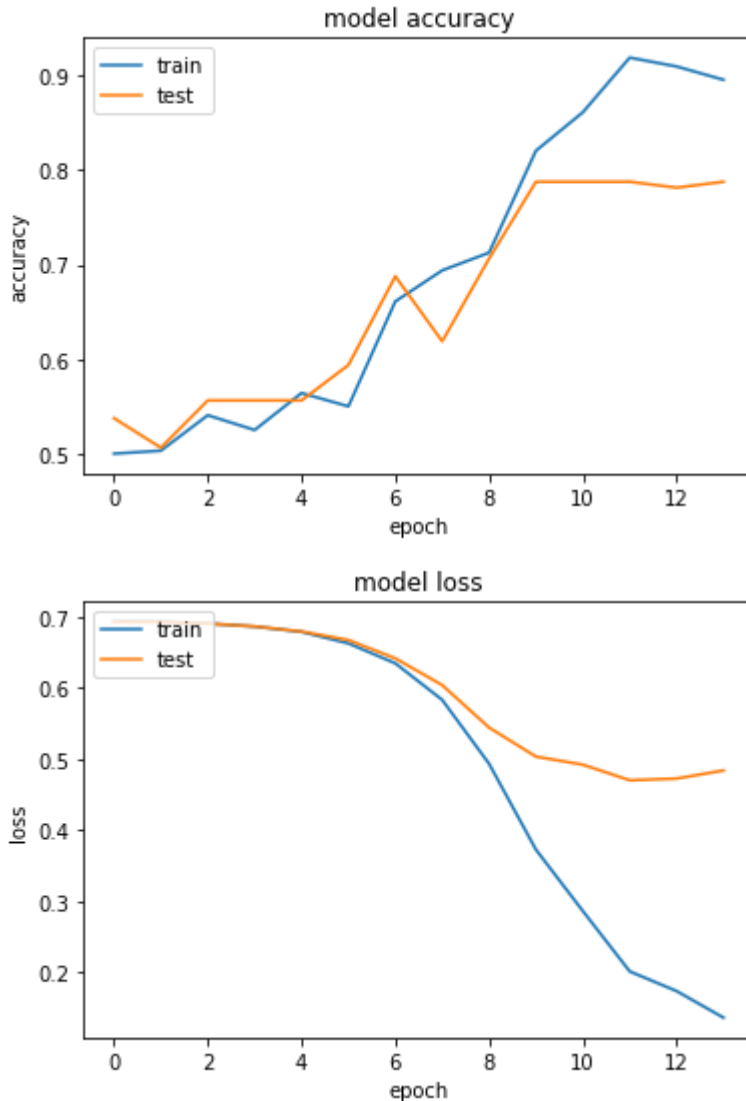
In [28]:

```
# Source: https://stackabuse.com/python-for-nlp-movie-sentiment-analysis-using-deep-Lea
plt.plot(history.history['acc'])
plt.plot(history.history['val_acc'])

plt.title('model accuracy')
plt.ylabel('accuracy')
plt.xlabel('epoch')
plt.legend(['train', 'test'], loc='upper left')
plt.show()
```

```
plt.plot(history.history['loss'])
plt.plot(history.history['val_loss'])

plt.title('model loss')
plt.ylabel('loss')
plt.xlabel('epoch')
plt.legend(['train', 'test'], loc='upper left')
plt.show()
```



D3. Assess the fitness of the model and any measures taken to address overfitting.

Rubric: The submission completely and accurately assesses the fitness of the model, and the assessment includes any measures taken to address overfitting.

Source: <https://towardsdatascience.com/dont-overfit-how-to-prevent-overfitting-in-your-deep-learning-models-63274e552323>

To decrease the complexity, I simply removed layers and reduced the number of neurons in order to make our network smaller. There is no general rule on how much to remove or how big your network should be.

At first the I had 4 Dense Layers and a GlobalAveragePoolin1D layer in my network. After viewing the results of the graph of the training process it was clear the model was over fitting. I decreased the number of nodes per Dense layer first and that helped but definitely did not solve the overfitting issue. After some more research on different types of network layers that are good for high accuracy results and a good fit I discovered the Bidirectional network layer with an input layer of LSTM. Updating the GolbalAveragePoolingLayer1D to the Bidirecitalon layer in addition to reducing the number of Dense layers and the total number of nodes for the remaining Dense layers the model had a much better fit.

D4. Discuss the predictive accuracy of the trained network.

Rubric: The submission discusses the predictive accuracy of the trained network using the selected evaluation metric from part D2.

The model has a predictive accuracy of 67%. This means that every 6 or 7 predictions from data input to the model will be correct.

In [29]:

```
print("Test Score:", score[0])  
print("Test Accuracy:", score[1])
```

Test Score: 0.4118857979774475
Test Accuracy: 0.824999988079071

Part V: Summary and Recommendations

E. Provide the code used to save the trained network within the neural network.

Rubric: The submission provides the code used to save the trained network within the neural network, and the code is complete and accurate.

F. Discuss the functionality of your neural network, including the impact of the network architecture.

Rubric: The submission accurately discusses the functionality of the neural network, including the impact of the network architecture. The discussion aligns with the research question from part A.

The neural network includes an input of ~2,200 customer reviews to train the model and ~550 reviews to test and validate the model. NLP was used to analyze the sentiments of the customers from the review where the actual sentiment label was 0 for a negative review and 1 for a positive review. The model was tuned to perform prediction on new customer reviews in order to identify customer sentiments if they were positive or negative reviews.

The impact of the network architecture was created through experimentation based on research. There are different keras layers that perform best for a NLP prediction model. Each layer's

parameters were adjusted until an acceptable degree of accuracy was reached to detect the sentiment of a review.

G. Recommend a course of action based on your results.

Rubric: The submission recommends an appropriate course of action based on the results as they relate to the research question.

The trained neural network is capable of processing natural language to identify sentiments from customer reviews. The neural network has a predictive accuracy of 82% and therefore capable of analyzing customer feedback and identifying sentiments in the review texts submitted by customers. This sentiment analysis can help the company to figure out which products are liked or not liked by customers based on the positive and negative reviews. Based on the analysis, the company can take appropriate action to address any concerns shown in the reviews.

Part VI: Reporting

H. Create your neural network using an industry-relevant interactive development environment (e.g., a Jupyter Notebook). Include a PDF or HTML document of your executed notebook presentation.

Rubric: The neural network is created in an industry-relevant interactive development environment and is complete, accurate, and in alignment with the data analysis of the report. A PDF or HTML document of the executed notebook presentation is provided.

I. List the web sources used to acquire data or segments of third-party code to support the application.

Rubric: The submission lists all web sources used to acquire data or segments of third-party code, and all the listed web sources are reliable.

1. <https://machinelearningmastery.com/use-word-embedding-layers-deep-learning-keras/>
2. <https://towardsdatascience.com/an-easy-tutorial-about-sentiment-analysis-with-deep-learning-and-keras-2bf52b9cba91>
3. <https://keras.io/api/callbacks/>
4. <https://machinelearningmastery.com/display-deep-learning-model-training-history-in-keras/>
5. <https://stackabuse.com/python-for-nlp-movie-sentiment-analysis-using-deep-learning-in-keras/>

J. Acknowledge sources, using in-text citations and references, for content that is quoted, paraphrased, or

summarized.

Rubric: The submission includes in-text citations for sources that are properly quoted, paraphrased, or summarized and a reference list that accurately identifies the author, date, title, and source location as available.

1. <https://keras.io/api/layers/activations/>
2. <https://stackoverflow.com/questions/34673396/what-does-the-standard-keras-model-output-mean-what-is-epoch-and-loss-in-keras#:~:text=Loss%3A%20A%20scalar%20value%20that,in%20Keras%2C%20Categorical%20Cross%20Entropy,loss%3A%20A%20scalar%20value%20that,in%20Keras%2C%20Categorical%20Cross%20Entropy>
3. <https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/>
4. <https://stats.stackexchange.com/questions/231061/how-to-use-early-stopping-properly-for-training-deep-neural-network>
5. https://keras.io/api/callbacks/early_stopping/
6. <https://towardsdatascience.com/dont-overfit-how-to-prevent-overfitting-in-your-deep-learning-models-63274e552323>
7. https://www.tensorflow.org/api_docs/python/tf/keras/layers/Embedding
8. <https://machinelearningmastery.com/use-word-embedding-layers-deep-learning-keras/>
9. https://tensorflow.org/api_docs/python/tf/keras/preprocessing/text/tokenizer

K. Demonstrate professional communication in the content and presentation of your submission.

Rubric: Content reflects attention to detail, is organized, and focuses on the main ideas as prescribed in the task or chosen by the candidate. Terminology is pertinent, is used correctly, and effectively conveys the intended meaning. Mechanics, usage, and grammar promote accurate interpretation and understanding.

