

```

import os
os.environ['TF_CPP_MIN_LOG_LEVEL'] = '2'
import numpy as np
import matplotlib.pyplot as plt

import json
import numpy as np
import keras.backend as K
from keras.utils import to_categorical
from keras.preprocessing.text import Tokenizer
from keras.preprocessing import sequence
from keras.models import Sequential
from keras.layers import Dense, Dropout, Embedding, LSTM,
Bidirectional

# 1. Loading the data
print("loading data...")

pos_file_name = "pos_amazon_cell_phone_reviews.json"
neg_file_name = "neg_amazon_cell_phone_reviews.json"
pos_file = open(pos_file_name, "r")
neg_file = open(neg_file_name, "r")
pos_data = json.loads(pos_file.read())['root']
neg_data = json.loads(neg_file.read())['root']
print("Positive data loaded. ", len(pos_data), "entries")
print("Negative data loaded. ", len(neg_data), "entries")

print("done loading data...")

labels = []
nlabels = []

# 2.Process reviews into sentences
pos_sentences, neg_sentences = [], []
for entry in pos_data :
    pos_sentences.append(entry['summary'] + " . " + entry['text'])
    labels.append(1)
for entry in neg_data :
    nlabels.append(0)
    neg_sentences.append(entry['summary'] + " . " + entry['text'])

print("No of Positive data found in AMAZON PRODUCT" )
print(len(pos_sentences))
print("No of Negative data found in AMAZON PRODUCT")
print(len(neg_sentences))

texts = pos_sentences + neg_sentences
labels = [1]*len(pos_sentences) + [0]*len(neg_sentences)

#print("after app", labels)

#print(type(pos_sentences), pos_sentences.shape, type(neg_sentences),
neg_sentences.shape)
#print(type(texts), texts.shape, type(labels), labels.shape)

```

```
# 3. Tokenize
tokenizer = Tokenizer()
tokenizer.fit_on_texts(texts)
sequences = tokenizer.texts_to_sequences(texts)

word_index = tokenizer.word_index
print('Found %s unique tokens.' % len(word_index))

# creating the dataset

a=10

data = {'Positive':len(pos_sentences), 'Negative':len(neg_sentences)}

courses = list(data.keys())
values = list(data.values())

fig = plt.figure(figsize = (5, 5))

# creating the bar plot
plt.bar(courses, values, color='blue',
        width = 0.4)

plt.ylabel("Number of reviews")
plt.title("Amazon Mobile phone reviews")
plt.show()

MAX_SEQUENCE_LENGTH = 50

data = sequence.pad_sequences(sequences, maxlen=MAX_SEQUENCE_LENGTH)
```