```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
# from wordcloud import WordCloud
import nltk
nltk.download("punkt")
nltk.download("wordnet")
nltk.download("stopwords")
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize

from nltk.stem import WordNetLemmatizer
from sklearn.model_selection import train_test_split

from tensorflow.keras.preprocessing.text import Tokenizer
from tensorflow.keras.preprocessing import sequence #unique id

from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, SimpleRNN, Dropout, Embedding

import warnings
warnings.filterwarnings("ignore")
```

Import "tensorflow.keras.preprocessing.text" could not be resolved (reportMissingImports)

Loading...

View Problem (Alt+F8)   No quick fixes available

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt.zip.
[nltk_data] Downloading package wordnet to /root/nltk_data...
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Unzipping corpora/stopwords.zip.
```

```python
df = pd.read_table('/content/amazon_alexa.tsv')
df.head()
```

	rating	date	variation	verified_reviews	feedback
0	5	31-Jul-18	Charcoal Fabric	Love my Echo!	1
1	5	31-Jul-18	Charcoal Fabric	Loved it!	1
2	4	31-Jul-18	Walnut Finish	Sometimes while playing a game, you can answer...	1
3	5	31-Jul-	Charcoal		1

```python
df.isnull().sum()
```

```
rating            0
date              0
variation         0
verified_reviews  0
feedback          0
dtype: int64
```

```python
df['variation'].value_counts()
```

```
Black  Dot                     516
Charcoal Fabric                430
Configuration: Fire TV Stick   350
Black  Plus                    270
Black  Show                    265
Black                          261
Black  Spot                    241
White  Dot                     184
Heather Gray Fabric            157
White  Spot                    109
```

```
White                          91
Sandstone Fabric               90
White   Show                   85
White   Plus                   78
Oak Finish                     14
Walnut Finish                   9
Name: variation, dtype: int64
```

```
len(df['variation'].value_counts())
```

Import "tensorflow.keras.preprocessing.text" could not be resolved (reportMissingImports)

Loading...

View Problem (Alt+F8)   No quick fixes available

```
    16
```

```python
def cleantext(text):
    tokens = word_tokenize(text.lower())
    ftoken = [t for t in tokens if(t.isalpha())]
    stop = stopwords.words("english")
    ctoken = [t for t in ftoken if(t not in stop)]
    lemma = WordNetLemmatizer()
    ltoken = [lemma.lemmatize(t) for t in ctoken]
    return " ".join(ltoken)
```

```python
df['verified_reviews']=df['verified_reviews'].apply(cleantext)
```

```python
sentlen = []
for sent in df["verified_reviews"]:
  sentlen.append(len(word_tokenize(sent)))
```

```python
df["SentLen"] = sentlen
df.head()
```

	rating	date	variation	verified_reviews	feedback	SentLen
0	5	31-Jul-18	Charcoal Fabric	love echo	1	2
1	5	31-Jul-18	Charcoal Fabric	loved	1	1
2	4	31-Jul-18	Walnut Finish	sometimes playing game answer question correct...	1	17
3	5	31-Jul-18	Charcoal Fabric	lot fun thing yr old learns dinosaur control l...	1	18
4	5	31-Jul-18	Charcoal Fabric	music	1	1

```python
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
x = df['verified_reviews']
y = le.fit_transform(df['variation'])
```

```python
np.quantile(sentlen, 0.95)
```

```
    40.0
```

```python
max_len = np.quantile(sentlen, 0.95)
max(df['SentLen'])
```

```
    245
```

```python
xtrain,xtest,ytrain,ytest = train_test_split(x,y,test_size=0.30,random_state=1)
```

```
tok = Tokenizer(char_level=False, split=" ")
tok.fit_on_texts(xtrain)


vocab_len = len(tok.index_word)
vocab_len
```

                 Import "tensorflow.keras.preprocessing.text" could not be
                 resolved (reportMissingImports)

        2945   Loading…

                 View Problem (Alt+F8)   No quick fixes available

```
seqtrain = tok.texts_to_sequences(xtrain) #step1
seqtrain
```

```
         112,
         528,
         349,
         199,
         128,
         961],
        [3, 17, 347, 152, 78, 352, 100, 936, 33, 35, 832],
        [8, 269],
        [207,
         2335,
         675,
         2336,
         44,
         259,
         2337,
         1483,
         908,
         2338,
         2339,
         417,
         2340,
         132,
         59,
         31,
         2341,
         2342,
         2343,
         29,
         520,
         20,
         29,
         822,
         501],
        [508, 17, 187, 1],
        [114,
         2344,
         424,
         1158,
         352,
         172,
         2345,
         95,
         334,
         70,
         720,
         1615,
         160,
         752,
         2346,
         2347,
         172,
         352],
        [16, 197, 122, 37],
        [107, 98, 576, 12, 577],
        [315, 149, 141],
        [1, 9, 36, 150, 256, 9],
        [2, 3, 479, 45, 127, 410],
        ...]
```

```
seqmattrain = sequence.pad_sequences(seqtrain, maxlen= int(max_len)) #step2
seqmattrain
```

```
array([[   0,    0,    0, ...,    8, 1231,  983],
       [   0,    0,    0, ...,    4,   92, 1232],
       [   0,    0,    0, ...,  203,  238,  278],
       ...,
       [   0,    0,    0, ...,  191,    3,   17],
       [   0,    0,    0, ...,    0,    0,  385],
       [   0,    0,    0, ...,    1,    2,   71]], dtype=int32)
```

Import "tensorflow.keras.preprocessing.text" could not be resolved (reportMissingImports)

Loading...

View Problem (Alt+F8)   No quick fixes available

```
seqtest = tok.texts_to_sequences(xtest)
seqmattest = sequence.pad_sequences(seqtest, maxlen=int(max_len))


rnn = Sequential()

rnn.add(Embedding(vocab_len+1,60, input_length=int(max_len), mask_zero=True))
rnn.add(SimpleRNN(units=32, activation="tanh"))
rnn.add(Dense(units=32, activation="relu"))
rnn.add(Dropout(0.2))

rnn.add(Dense(units=1, activation="sigmoid"))

rnn.compile(optimizer="adam", loss="binary_crossentropy")

rnn.fit(seqmattrain, ytrain, batch_size=50, epochs=50)

ypred = rnn.predict(seqmattest)

ypred = ypred>0.5
```

```
Epoch 41/50
45/45 [==============================] - 1s 17ms/step - loss: -17504.9453
Epoch 42/50
45/45 [==============================] - 1s 30ms/step - loss: -18310.8770
Epoch 43/50
45/45 [==============================] - 1s 31ms/step - loss: -19923.4668
Epoch 44/50
45/45 [==============================] - 1s 21ms/step - loss: -19898.6230
Epoch 45/50
45/45 [==============================] - 1s 18ms/step - loss: -20831.3770
Epoch 46/50
45/45 [==============================] - 1s 18ms/step - loss: -21624.8281
Epoch 47/50
45/45 [==============================] - 1s 18ms/step - loss: -22551.4258
Epoch 48/50
45/45 [==============================] - 1s 18ms/step - loss: -23493.5605
Epoch 49/50
45/45 [==============================] - 1s 18ms/step - loss: -24575.0801
Epoch 50/50
45/45 [==============================] - 1s 18ms/step - loss: -25301.2090
30/30 [==============================] - 0s 4ms/step
```

Import "tensorflow.keras.preprocessing.text" could not be resolved (reportMissingImports)

Loading...

View Problem (Alt+F8)   No quick fixes available

```
from sklearn.metrics import classification_report
print(classification_report(ytest,ypred))
```

```
              precision    recall  f1-score   support

           0       0.25      0.04      0.07        75
           1       0.15      0.99      0.26       142
           2       0.00      0.00      0.00        76
           3       0.00      0.00      0.00        97
           4       0.00      0.00      0.00        75
           5       0.00      0.00      0.00       138
           6       0.00      0.00      0.00        98
           7       0.00      0.00      0.00        48
           8       0.00      0.00      0.00         5
           9       0.00      0.00      0.00        33
          10       0.00      0.00      0.00         2
          11       0.00      0.00      0.00        32
          12       0.00      0.00      0.00        47
          13       0.00      0.00      0.00        20
          14       0.00      0.00      0.00        27
          15       0.00      0.00      0.00        30

    accuracy                           0.15       945
   macro avg       0.03      0.06      0.02       945
weighted avg       0.04      0.15      0.04       945
```

✓ 0s   completed at 12:29 AM   ● ✕

Import "tensorflow.keras.preprocessing.text" could not be resolved (reportMissingImports)

Loading...

View Problem (Alt+F8)   No quick fixes available