⌄ Fake News Classifier Using LSTM RNN

```python
import pandas as pd
import numpy as np
import os
from pathlib import Path
from zipfile import ZipFile
import json
```

```python
!pip install kaggle
```

```
Requirement already satisfied: kaggle in /usr/local/lib/python3.10/dist-packages (1.6.12)
Requirement already satisfied: six>=1.10 in /usr/local/lib/python3.10/dist-packages (from kaggle) (1.16.0)
Requirement already satisfied: certifi>=2023.7.22 in /usr/local/lib/python3.10/dist-packages (from kaggle) (2024.2.2)
Requirement already satisfied: python-dateutil in /usr/local/lib/python3.10/dist-packages (from kaggle) (2.8.2)
Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-packages (from kaggle) (2.31.0)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages (from kaggle) (4.66.2)
Requirement already satisfied: python-slugify in /usr/local/lib/python3.10/dist-packages (from kaggle) (8.0.4)
Requirement already satisfied: urllib3 in /usr/local/lib/python3.10/dist-packages (from kaggle) (2.0.7)
Requirement already satisfied: bleach in /usr/local/lib/python3.10/dist-packages (from kaggle) (6.1.0)
Requirement already satisfied: webencodings in /usr/local/lib/python3.10/dist-packages (from bleach->kaggle) (0.5.1)
Requirement already satisfied: text-unidecode>=1.3 in /usr/local/lib/python3.10/dist-packages (from python-slugify->kaggle) (1.3)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests->kaggle) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests->kaggle) (3.7)
```

```python
with ZipFile('/content/train.csv.zip','r') as zipfile:
  zipfile.extractall()
```

```python
!ls
```

```
sample_data   train.csv   train.csv.zip
```

```python
df = pd.read_csv('/content/train.csv')
```

```python
df.head()
```

	id	title	author	text	label
0	0	House Dem Aide: We Didn't Even See Comey's Let...	Darrell Lucus	House Dem Aide: We Didn't Even See Comey's Let...	1
1	1	FLYNN: Hillary Clinton, Big Woman on Campus - ...	Daniel J. Flynn	Ever get the feeling your life circles the rou...	0
2	2	Why the Truth Might Get You Fired	Consortiumnews.com	Why the Truth Might Get You Fired October 29, ...	1
3	3	15 Civilians Killed In Single US		Videos 15 Civilians Killed In	

```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20800 entries, 0 to 20799
Data columns (total 5 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   id      20800 non-null  int64
 1   title   20242 non-null  object
 2   author  18843 non-null  object
 3   text    20761 non-null  object
 4   label   20800 non-null  int64
dtypes: int64(2), object(3)
memory usage: 812.6+ KB
```

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

```
id          0
title     558
author   1957
text       39
label       0
dtype: int64
```

∨   We will take the title of all the data and train the LSTM RNN to find out uf it is fake or not

```python
df = df.dropna()
```

```python
X = df.drop('label',axis=1)
y = df['label']
```

```python
X.shape
```

```
(18285, 4)
```

```python
y.shape
```

```
(18285,)
```

```python
!pip install tensorflow
```

```
Requirement already satisfied: tensorflow in /usr/local/lib/python3.10/dist-packages (2.15.0)
Requirement already satisfied: absl-py>=1.0.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (1.4.0)
Requirement already satisfied: astunparse>=1.6.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (1.6.3)
Requirement already satisfied: flatbuffers>=23.5.26 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (24.3.25)
Requirement already satisfied: gast!=0.5.0,!=0.5.1,!=0.5.2,>=0.2.1 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (0.5.4)
Requirement already satisfied: google-pasta>=0.1.1 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (0.2.0)
Requirement already satisfied: h5py>=2.9.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (3.9.0)
Requirement already satisfied: libclang>=13.0.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (18.1.1)
Requirement already satisfied: ml-dtypes~=0.2.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (0.2.0)
Requirement already satisfied: numpy<2.0.0,>=1.23.5 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (1.25.2)
Requirement already satisfied: opt-einsum>=2.3.2 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (3.3.0)
Requirement already satisfied: packaging in /usr/local/lib/python3.10/dist-packages (from tensorflow) (24.0)
Requirement already satisfied: protobuf!=4.21.0,!=4.21.1,!=4.21.2,!=4.21.3,!=4.21.4,!=4.21.5,<5.0.0dev,>=3.20.3 in /usr/local/lib/python
Requirement already satisfied: setuptools in /usr/local/lib/python3.10/dist-packages (from tensorflow) (67.7.2)
Requirement already satisfied: six>=1.12.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (1.16.0)
Requirement already satisfied: termcolor>=1.1.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (2.4.0)
Requirement already satisfied: typing-extensions>=3.6.6 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (4.11.0)
Requirement already satisfied: wrapt<1.15,>=1.11.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (1.14.1)
Requirement already satisfied: tensorflow-io-gcs-filesystem>=0.23.1 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (0.37.0
Requirement already satisfied: grpcio<2.0,>=1.24.3 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (1.63.0)
Requirement already satisfied: tensorboard<2.16,>=2.15 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (2.15.2)
Requirement already satisfied: tensorflow-estimator<2.16,>=2.15.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (2.15.0)
Requirement already satisfied: keras<2.16,>=2.15.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (2.15.0)
Requirement already satisfied: wheel<1.0,>=0.23.0 in /usr/local/lib/python3.10/dist-packages (from astunparse>=1.6.0->tensorflow) (0.43.
Requirement already satisfied: google-auth<3,>=1.6.3 in /usr/local/lib/python3.10/dist-packages (from tensorboard<2.16,>=2.15->tensorflc
Requirement already satisfied: google-auth-oauthlib<2,>=0.5 in /usr/local/lib/python3.10/dist-packages (from tensorboard<2.16,>=2.15->te
Requirement already satisfied: markdown>=2.6.8 in /usr/local/lib/python3.10/dist-packages (from tensorboard<2.16,>=2.15->tensorflow) (3.
Requirement already satisfied: requests<3,>=2.21.0 in /usr/local/lib/python3.10/dist-packages (from tensorboard<2.16,>=2.15->tensorflow)
Requirement already satisfied: tensorboard-data-server<0.8.0,>=0.7.0 in /usr/local/lib/python3.10/dist-packages (from tensorboard<2.16,>
Requirement already satisfied: werkzeug>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from tensorboard<2.16,>=2.15->tensorflow) (3.
Requirement already satisfied: cachetools<6.0,>=2.0.0 in /usr/local/lib/python3.10/dist-packages (from google-auth<3,>=1.6.3->tensorboar
Requirement already satisfied: pyasn1-modules>=0.2.1 in /usr/local/lib/python3.10/dist-packages (from google-auth<3,>=1.6.3->tensorboard
Requirement already satisfied: rsa<5,>=3.1.4 in /usr/local/lib/python3.10/dist-packages (from google-auth<3,>=1.6.3->tensorboard<2.16,>=
Requirement already satisfied: requests-oauthlib>=0.7.0 in /usr/local/lib/python3.10/dist-packages (from google-auth-oauthlib<2,>=0.5->t
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=2.21.0->tensorboar
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=2.21.0->tensorboard<2.16,>=2.1
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=2.21.0->tensorboard<2.16
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=2.21.0->tensorboard<2.16
Requirement already satisfied: MarkupSafe>=2.1.1 in /usr/local/lib/python3.10/dist-packages (from werkzeug>=1.0.1->tensorboard<2.16,>=2.
Requirement already satisfied: pyasn1<0.7.0,>=0.4.6 in /usr/local/lib/python3.10/dist-packages (from pyasn1-modules>=0.2.1->google-auth<
Requirement already satisfied: oauthlib>=3.0.0 in /usr/local/lib/python3.10/dist-packages (from requests-oauthlib>=0.7.0->google-auth-oa
```

```python
import tensorflow as tf
```

```python
from tensorflow.keras.preprocessing.text import one_hot
from tensorflow.keras.preprocessing.sequence import pad_sequences
from tensorflow.keras.layers import Embedding
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import LSTM
from tensorflow.keras.layers import Dense
```

```python
## Setting vocabulary size

voc_size = 5000
```

⌄ One Hot Representation

```
messages = X.copy()
```

```
messages.reset_index(inplace=True)
```

```
messages
```

	index	id	title	author	text
0	0	0	House Dem Aide: We Didn't Even See Comey's Let...	Darrell Lucus	House Dem Aide: We Didn't Even See Comey's Let...
1	1	1	FLYNN: Hillary Clinton, Big Woman on Campus - ...	Daniel J. Flynn	Ever get the feeling your life circles the rou...
2	2	2	Why the Truth Might Get You Fired	Consortiumnews.com	Why the Truth Might Get You Fired October 29, ...
3	3	3	15 Civilians Killed In Single US Airstrike Hav...	Jessica Purkiss	Videos 15 Civilians Killed In Single US Airstr...
4	4	4	Iranian woman jailed for fictional unpublished...	Howard Portnoy	Print \nAn Iranian woman has been sentenced to...
...	...	...	...	...	...
			Rapper T.I.: Trump a		Rapper T. I. unloaded

```
import nltk
import re
from nltk.corpus import stopwords
```

```
nltk.download('stopwords')
```

```
    [nltk_data] Downloading package stopwords to /root/nltk_data...
    [nltk_data]   Unzipping corpora/stopwords.zip.
    True
```

```
## Data preprocessing

from nltk.stem.porter import PorterStemmer
ps=PorterStemmer()
corpus=[]

for i in range(0,len(messages)):
  review  = re.sub('[^a-zA-Z]', ' ' ,messages['title'][i])
  review = review.lower()
  review  = review.split()

  review = [ps.stem(word) for word in review if not word in stopwords.words('english')]
  review = ' '.join(review)
  corpus.append(review)


corpus


corpus[1]
```

```
    'flynn hillari clinton big woman campu breitbart'
```

```
### One - Hot Representation

onehot_repr = [one_hot(word,voc_size) for word in corpus ]
onehot_repr


onehot_repr[1]
```

```
    [1461, 2920, 2231, 744, 561, 2654, 4745]
```

⌄ Embedding layer or Word to Vector

```
### Padding to create an equal length input


sent_length = 20

embedded_docs = pad_sequences(onehot_repr,padding='pre',maxlen=sent_length)
embedded_docs
```

```
    array([[   0,    0,    0, ...,  424,  494, 2641],
           [   0,    0,    0, ...,  561, 2654, 4745],
           [   0,    0,    0, ..., 3025, 4385, 3666],
           ...,
           [   0,    0,    0, ..., 2052, 2218, 2794],
           [   0,    0,    0, ..., 1648, 2621, 4346],
           [   0,    0,    0, ..., 1852, 4872, 3868]], dtype=int32)
```

```
embedded_docs[1]
```

```
    array([   0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
              0,    0, 1461, 2920, 2231,  744,  561, 2654, 4745], dtype=int32)
```

```
from tensorflow.keras.layers import Dropout


### Creating model

embedding_vector_feature = 40 #each and every word is going to have 40 fearures
model = Sequential()
model.add(Embedding(voc_size,embedding_vector_feature,input_length=sent_length))
model.add(Dropout(0.3))
model.add(LSTM(100))
model.add(Dropout(0.3))
model.add(Dense(1,activation='sigmoid'))
model.compile(loss='binary_crossentropy',optimizer='adam',metrics=['accuracy'])


print(model.summary())
```

```
    Model: "sequential_3"
```

```
Layer (type)                Output Shape              Param #
=================================================================
embedding_3 (Embedding)     (None, 20, 40)            200000

dropout (Dropout)           (None, 20, 40)            0

lstm_2 (LSTM)               (None, 100)               56400

dropout_1 (Dropout)         (None, 100)               0

dense_2 (Dense)             (None, 1)                 101

=================================================================
Total params: 256501 (1001.96 KB)
Trainable params: 256501 (1001.96 KB)
Non-trainable params: 0 (0.00 Byte)
_____
None
```

```python
import numpy as np
```

```python
X_final = np.array(embedded_docs)
```

```python
y_final=np.array(y)
```

```python
X_final.shape
```

```
(18285, 20)
```

```python
y_final.shape
```

```
(18285,)
```

```python
y_final
```

```
array([1, 0, 1, ..., 0, 1, 1])
```

Train Test Split

```python
from sklearn.model_selection import train_test_split
```

```python
X_train,X_test,y_train,y_test = train_test_split(X_final,y_final,test_size=0.30,random_state=0)
```

⌄ Model Training

```python
model.fit(X_train,y_train,validation_data=(X_test,y_test),epochs=10,batch_size=64)
```

```
Epoch 1/10
200/200 [==============================] - 14s 51ms/step - loss: 0.3480 - accuracy: 0.8346 - val_loss: 0.1990 - val_accuracy: 0.9149
Epoch 2/10
200/200 [==============================] - 12s 62ms/step - loss: 0.1519 - accuracy: 0.9402 - val_loss: 0.1847 - val_accuracy: 0.9200
Epoch 3/10
200/200 [==============================] - 13s 67ms/step - loss: 0.1163 - accuracy: 0.9564 - val_loss: 0.1982 - val_accuracy: 0.9176
Epoch 4/10
200/200 [==============================] - 15s 77ms/step - loss: 0.0920 - accuracy: 0.9669 - val_loss: 0.2205 - val_accuracy: 0.9165
Epoch 5/10
200/200 [==============================] - 18s 88ms/step - loss: 0.0735 - accuracy: 0.9741 - val_loss: 0.2899 - val_accuracy: 0.8994
Epoch 6/10
200/200 [==============================] - 21s 106ms/step - loss: 0.0589 - accuracy: 0.9798 - val_loss: 0.2530 - val_accuracy: 0.9085
Epoch 7/10
200/200 [==============================] - 11s 54ms/step - loss: 0.0453 - accuracy: 0.9849 - val_loss: 0.3182 - val_accuracy: 0.9149
Epoch 8/10
200/200 [==============================] - 9s 47ms/step - loss: 0.0359 - accuracy: 0.9877 - val_loss: 0.3654 - val_accuracy: 0.9125
Epoch 9/10
200/200 [==============================] - 11s 56ms/step - loss: 0.0251 - accuracy: 0.9916 - val_loss: 0.3785 - val_accuracy: 0.9156
Epoch 10/10
200/200 [==============================] - 11s 54ms/step - loss: 0.0206 - accuracy: 0.9941 - val_loss: 0.3753 - val_accuracy: 0.9152
<keras.src.callbacks.History at 0x7b5a73f4a6e0>
```

PERFORMANCE METRICS

```python
y_pred = model.predict(X_test)
```

```
172/172 [==============================] - 3s 8ms/step
```

```python
y_pred=np.where(y_pred >0.5,1,0)   #whenever my y_pred is greater than 0.5, then put it as 1 or fake
```

```python
from sklearn.metrics import accuracy_score, confusion_matrix,classification_report
```

```python
print(confusion_matrix(y_test,y_pred))
```

```
[[2810  274]
 [ 191 2211]]
```

```python
print(accuracy_score(y_test,y_pred))
```

```
0.9152387896463726
```

```python
print(classification_report(y_test,y_pred))
```

```
              precision    recall  f1-score   support

           0       0.94      0.91      0.92      3084
           1       0.89      0.92      0.90      2402

    accuracy                           0.92      5486
   macro avg       0.91      0.92      0.91      5486
weighted avg       0.92      0.92      0.92      5486
```

Start coding or generate with AI.