E:\spam.py

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

import string
import nltk
from nltk.corpus import stopwords
from wordcloud import WordCloud
nltk.download('stopwords')

# Importing libraries necessary for Model Building and Training
import tensorflow as tf
from tensorflow.keras.preprocessing.text import Tokenizer
from tensorflow.keras.preprocessing.sequence import pad_sequences
from sklearn.model_selection import train_test_split
from keras.callbacks import EarlyStopping, ReduceLROnPlateau

import warnings
warnings.filterwarnings('ignore')
data = pd.read_csv('Emails.csv')
data.head()
data.shape
sns.countplot(x='spam', data=data)
plt.show()
# Downsampling to balance the dataset
ham_msg = data[data.spam == 0]
spam_msg = data[data.spam == 1]
ham_msg = ham_msg.sample(n=len(spam_msg),random_state=42)

# Plotting the counts of down sampled dataset
balanced_data = ham_msg.append(spam_msg).reset_index(drop=True)
plt.figure(figsize=(8, 6))
sns.countplot(data = balanced_data, x='spam')
plt.title('Distribution of Ham and Spam email messages after downsampling')
plt.xlabel('Message types')
balanced_data['text'] = balanced_data['text'].str.replace('Subject', '')
balanced_data.head()
punctuations_list = string.punctuation
def remove_punctuations(text):
temp = str.maketrans('', '', punctuations_list)
return text.translate(temp)

balanced_data['text']= balanced_data['text'].apply(lambda x: remove_punctuations(x))
balanced_data.head()
def plot_word_cloud(data, typ):
email_corpus = " ".join(data['text'])
plt.figure(figsize=(7, 7))
wc = WordCloud(background_color='black',max_words=100,width=800,height=400,collocations=False)
.generate(email_corpus)
plt.imshow(wc, interpolation='bilinear')
plt.title(f'WordCloud for {typ} emails', fontsize=15)
plt.axis('off')
plt.show()
```

```
54 plot_word_cloud(balanced_data[balanced_data['spam'] == 0], typ='Non-Spam')
55 plot_word_cloud(balanced_data[balanced_data['spam'] == 1], typ='Spam')
56
```