

ASSIGNMENT 4

Problem Statement :- SMS SPAM Classification

Assignment Date	03 NOVEMBER 2022
Student Name	Sheela iswaryam.k
Student Roll Number	PNT2022TMTD36873
Maximum Marks	2 MARKS

1. Download The Dataset :

```
1: from google.colab import drive
   drive.mount('/content/drive')
```

Mounted at /content/drive

2. Import required library

```
[2]: import csv
      import tensorflow as tf
      import pandas as pd
      import numpy as np
      import seaborn as sns
      import re
      import matplotlib.pyplot as plt
```

```
In [3]: from tensorflow.keras.preprocessing.text import Tokenizer
      from tensorflow.keras.preprocessing.text import one_hot
      from tensorflow.keras.preprocessing.sequence import pad_sequences
      from tensorflow.keras.models import Sequential
      from tensorflow.keras.layers import LSTM
      from tensorflow.keras.layers import Dense
      from tensorflow.keras.layers import Embedding
      from tensorflow.keras.optimizers import Adam
      from sklearn.preprocessing import LabelEncoder
      from sklearn.model_selection import train_test_split
      from sklearn.metrics import accuracy_score, confusion_matrix
```

```
40: [5]: import nltk
      nltk.download('stopwords')
      from nltk.corpus import stopwords
      from nltk.stem import PorterStemmer
      STOPWORDS = set(stopwords.words('english'))

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

3. Read dataset and do pre-processing

```
In [5]: data=pd.read_csv("/content/drive/MyDrive/spam.csv",encoding="latin")
data.head()
```

```
Out[5]:
```

	v1	v2	Unnamed: 2	Unnamed: 3	Unnamed: 4
0	ham	Go until jurong point, crazy.. Available only ...	NaN	NaN	NaN
1	ham	Ok lar... Joking wif u oni...	NaN	NaN	NaN
2	spam	Free entry in 2 a wkly comp to win FA Cup fina...	NaN	NaN	NaN
3	ham	U dun say so early hor... U c already then say...	NaN	NaN	NaN
4	ham	Nah I don't think he goes to usf, he lives aro...	NaN	NaN	NaN

```
In [6]: data.tail()
```

```
Out[6]:
```

	v1	v2	Unnamed: 2	Unnamed: 3	Unnamed: 4
5567	spam	This is the 2nd time we have tried 2 contact u...	NaN	NaN	NaN
5568	ham	Will l_ b going to esplanade fr home?	NaN	NaN	NaN
5569	ham	Pity, * was in mood for that. So...any other s...	NaN	NaN	NaN
5570	ham	The guy did some bitching but I acted like i'd...	NaN	NaN	NaN
5571	ham	Rofl. Its true to its name	NaN	NaN	NaN

```
In [7]: data=data.drop(columns=["Unnamed: 2","Unnamed: 3","Unnamed: 4"])
```

```
In [9]: data=data.rename({"v1":"Category","v2":"Message"},axis=1)
data.isnull().sum()
```

```
Out[9]: Category    0
Message    0
dtype: int64
```

```
In [10]: data["Message Length"]=data["Message"].apply(len)
```

```
In [11]: fig=plt.figure(figsize=(12,8))
sns.histplot(
    x=data["Message Length"],
    hue=data["Category"]
)
plt.title("ham & spam messege length comparision")
plt.show()
```

PDF Converter

Only two pages were converted.

Please **Sign Up** to convert the full document.

www.freepdfconvert.com/membership