

## Assignment -4

### Problem Statement :- SMS SPAM Classification

Assignment Date	20 September 2022
Student name	SHANMUGAM M
Student roll number	<b>112519106004</b>
Maximum marks	2 marks

```

##IMPORT LIBRARIES
import pandas as pd

import numpy as np

import nltk import
re

nltk.download('stopwords')

from nltk.corpus import stopwords from
nltk.stem.porter import PorterStemmer

##LOAD DATASET
a = pd.read_csv('/content/spam.csv',encoding='ISO-8859-1')
a.head()
a=a[['v1','v2']]
a.head()
a.shape

##Text processing (NLP)
ps=PorterStemmer()
message=[] for i in
range(0,5572):
    msg=a['v2'][i] msg=re.sub('[^a-zA-
Z]', '',msg) msg=msg.lower()

    msg=msg.split(' ') msg = [ps.stem(word) for word in msg if word not in
set(stopwords.words('english'))] msg=' '.join(msg) message.append(msg)

```

```
message[:6] from sklearn.feature_extraction.text import
```

```
CountVectorizer cv = CountVectorizer() x =
```

```
cv.fit_transform(message).toarray()
```

```
x
```

```
#LABEL ENCODING
```

```
from sklearn.preprocessing import LabelEncoder le
```

```
= LabelEncoder()
```

```
a['v1']=le.fit_transform(a['v1'])
```

```
y = a['v1'].values y
```

```
##MODEL BUILDIND
```

```
from tensorflow.keras.models import Sequential from
```

```
tensorflow.keras.layers import Dense
```

```
model = Sequential() model.add(Dense(1550,activation='relu'))
```

```
model.add(Dense(3000,activation='relu'))
```

```
model.add(Dense(1,activation='sigmoid'))
```

```
model.compile(optimizer='adam',loss='binary_crossentropy',metrics=['accuracy'])
```

```
model.fit(x,y,epochs=10)
```

```
##SAVE THE MODEL
```

```
model.save('spam-NLP.h5') ##TEST
```

```
THE MODEL
```

```
msg='FREE MESSAGE Activate your 500 FREE Text Messages by replying to this message with the word
FREE' print('THE ORIGINAL MESSAGE IS: ',msg) msg=re.sub('[^a-zA-Z]',' ',msg) msg=msg.lower()
msg=msg.split(' ') msg = [ps.stem(word) for word in msg if word not in
set(stopwords.words('english'))] msg=' '.join(msg) print('THE STEMMED MESSAGE IS:
',msg)
```

```
predict = model.predict(cv.transform([msg])) if predict > 0.5: pred='SPAM' else:
pred='NOT SPAM' print('THE MESSAGE IS PREDICTED AS: ',pred) msg='Wishing you
and your family Merry \X\" mas and HAPPY NEW Year in advance.."' print('THE
ORIGINAL MESSAGE IS: ',msg) msg=re.sub('[^a-zA-Z]',' ',msg) msg=msg.lower()
msg=msg.split(' ') msg = [ps.stem(word) for word in msg if word not in
set(stopwords.words('english'))] msg=' '.join(msg) print('THE STEMMED MESSAGE IS:
',msg)
```

```
predict = model.predict(cv.transform([msg]))
if predict > 0.5: pred='spam'
else: pred='NOT SPAM'
print('THE MESSAGE IS PREDICTED AS: ',pred)
```