Date	18 October 2023
Team ID	NM2023TMID335
Team Name	Proj_227279_Team_1
Project Name	Building a smarter Ai-powered spam classifier

Phase 3: Development Part 1

In this phase, start building the core components of our spam classifier. you will begin building your project by loading and preprocessing the dataset.

1   Data Preprocessing.

```python
[1]: # Import the libraries.
     import pandas as pd
     import numpy as np
     import matplotlib.pyplot as plt
     import seaborn as sns

     import os
     print(os.listdir(r"C:\Users\NELSON\Desktop\new deploy\sanjay"))
```

```
['.ipynb_checkpoints', 'Datst.csv', 'final.xlsx', 'prep.ipynb',
'preprocessing.ipynb',       'sanjay1.csv',       'sanjay2.csv',
'spamclassifier.csv']
```

```python
[2]: # Read the Dataset.
     data = pd.read_csv(r"C:\Users\NELSON\Desktop\new deploy\sanjay/spamclassifier.
       csv", encoding="latin-1")
```

```python
[3]: import warnings
     warnings.filterwarnings('ignore')
```

```python
[4]: data.head()
```

```
[4]:        v1                                              v2 Unnamed: 2  \
0   ham Go until jurong point, crazy.. Available only …NaN

1   ham    Ok lar… Joking wif u oni…   NaN

2   spam Free entry in 2 a wkly comp to win FA Cup fina…     NaN

3   ham U dun say so early hor… U c already then say…  NaN

4   ham Nah I don't think he goes to usf, he lives aro…NaN

Unnamed: 3 Unnamed: 4

0        NaNNaN

1        NaNNaN

2        NaNNaN
```

```
3        NaNNaN

4        NaNNaN
```

[5]: `df = data.dropna()`

[6]: `df.columns`

[6]: `Index(['v1', 'v2', 'Unnamed: 2', 'Unnamed: 3', 'Unnamed: 4'], dtype='object')`

[7]: `# Describe the data in statastical view.`
`df.describe()`

[7]:
```
       v1   v2 Unnamed: 2 \ count 6      6 6 unique 1       5 5
top ham Edison has rightly said, \A fool can ask more … GN freq 6
2 2


        Unnamed: 3 Unnamed: 4
count          6          6

unique         5          5

top           GE   GNT:-)"

freq           2          2
```

[8]: `df.info()`

```
<class 'pandas.core.frame.DataFrame'>
Index: 6 entries, 281 to 5048
Data    columns    (total    5
columns):
 # Column Non-Null Count Dtype --- -
----- -------------- -----
```
```
0    v1    6 non-null object

1    v2    6 non-null object

2    Unnamed: 2 6 non-null object

3    Unnamed: 3 6 non-null object 4 Unnamed: 4 6 non-null object

dtypes: object(5)

memory usage: 288.0+ bytes
```

[9]: `df.duplicated()`

[9]: `281    False`

[10]: `df.size`

```
        1038    False
        2255    False
        3525    False
        4668    False
        5048     True
        dtype: bool
```

[10]: 30

[11]: `df.shape`

[11]: (6, 5)

[12]: `df.tail()`

[12]:
```
           v1                                          v2  \
1038 ham Edison has rightly said, \A fool can ask more …
2255 ham  I just lov this line: \Hurt me with the truth
3525 ham \HEY BABE! FAR 2 SPUN-OUT 2 SPK AT DA MO… DE…
4668 ham When I was born, GOD said, \Oh No! Another IDI…
5048 ham Edison has rightly said, \A fool can ask more …

              Unnamed: 2                             Unnamed: 3  \
1038                 GN                                     GE
2255   I don't mind i wil tolerat.bcs ur my someone… But
3525  HAD A COOL NYTHO                       TX 4 FONIN HON
4668        GOD said         \"OH No! COMPETITION\". Who
                                                       knew
5048                 GN                                     GE


                                   Unnamed: 4
1038                                  GNT:-)"

2255Never comfort me with a lie\" gud ni8 and swe…
3525          CALL 2MWEN IM BK FRMCLOUD 9! J X\""
4668one day these two will become FREINDS FOREVER!"
5048                                  GNT:-)"
```