

In this project, we will use Python to create an email spam detection system. Our goal is to train a machine learning model that can classify emails as either spam or non-spam (ham). This will help us filter out unwanted and potentially harmful emails from our inbox. We'll follow the standard data science workflow, including data loading, preprocessing, feature extraction, model training, evaluation, and prediction. Let's begin building our email spam detector!

BUILDING A SMARTER AI-POWERED SPAM CLASSIFIER



UNDERSTANDING A DATA

```
[ ] #importing libraries

import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, confusion_matrix, roc_curve, roc_auc_score
import nltk
from nltk.corpus import stopwords
from collections import Counter

#libraries for data visualization
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

```
[ ] df= pd.read_csv("/content/spam.csv",encoding='ISO-8859-1')
df
```

	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
...	...	...	...	...	...
5567	spam	This is the 2nd time we have tried 2 contact u...	NaN	NaN	NaN
5568	ham	Will i_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

5572 rows x 5 columns

```
[ ] df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5572 entries, 0 to 5571
Data columns (total 5 columns):
#   Column          Non-Null Count  Dtype
---  -
0   v1              5572 non-null   object
1   v2              5572 non-null   object
2   Unnamed: 2      50 non-null     object
3   Unnamed: 3      12 non-null     object
4   Unnamed: 4      6 non-null      object
dtypes: object(5)
memory usage: 217.8+ KB
```

```
[ ] # Downloading the stopwords dataset

nltk.download('stopwords')

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

[ ] # Drop unnecessary columns from the DataFrame

columns_to_drop = ["Unnamed: 2", "Unnamed: 3", "Unnamed: 4"]
df.drop(columns=columns_to_drop, inplace=True)
```

```
[ ] df
```

```

      v1      v2
0  ham  Go until jurong point, crazy.. Available only ...
1  ham  Ok lar... Joking wif u oni...
2  spam  Free entry in 2 a wkly comp to win FA Cup fina...
3  ham  U dun say so early hor... U c already then say...
4  ham  Nah I don't think he goes to usf, he lives aro...
...  ...  ...
5567 spam  This is the 2nd time we have tried 2 contact u...
5568 ham  Will i_b going to esplanade fr home?
5569 ham  Pity, * was in mood for that. So...any other s...
5570 ham  The guy did some bitching but I acted like i'd...
5571 ham  Rofl. Its true to its name

5572 rows x 2 columns
```

```
[ ] # Rename the columns "v1" and "v2" to new names

new_column_names = {"v1": "Category", "v2": "Message"}
df.rename(columns = new_column_names, inplace = True)
```

```
[ ] df[df.duplicated()]
```

```

      Category      Message
102      ham  As per your request 'Melle Melle (Oru Minnamin...
153      ham  As per your request 'Melle Melle (Oru Minnamin...
206      ham  As I entered my cabin my PA said, " Happy B'd...
222      ham  Sorry, I'll call later
325      ham  No calls..messages..missed calls
...      ...  ...
5524     spam  You are awarded a SiPix Digital Camera! call 0...
5535      ham  I know you are thinkin malaria. But relax, chi...
5539      ham  Just sleeping..and surfing
5553      ham  Hahaha..use your brain dear
5558      ham  Sorry, I'll call later

403 rows x 2 columns
```

```
[ ] #Drop duplicated values
df=df.drop_duplicates()
df
```

```

      Category      Message
0      ham  Go until jurong point, crazy.. Available only ...
1      ham  Ok lar... Joking wif u oni...
2      spam  Free entry in 2 a wkly comp to win FA Cup fina...
3      ham  U dun say so early hor... U c already then say...
4      ham  Nah I don't think he goes to usf, he lives aro...
...      ...  ...
5567     spam  This is the 2nd time we have tried 2 contact u...
5568      ham  Will i_b going to esplanade fr home?
5569      ham  Pity, * was in mood for that. So...any other s...
5570      ham  The guy did some bitching but I acted like i'd
```