In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.linear_model import LogisticRegression
```

In [2]:
```
df=pd.read_csv(r"C:\Users\user\Downloads\bot detection.csv")
df
```

Out[2]:

	User ID	Username	Tweet	Retweet Count	Mention Count	Follower Count	Verified	Bot Label	Locatio
0	132131	flong	Station activity person against natural majori...	85	1	2353	False	1	Adkinsto
1	289683	hinesstephanie	Authority research natural life material staff...	55	5	9617	True	0	Sandersto
2	779715	roberttran	Manage whose quickly especially foot none to g...	6	2	4363	True	0	Harrisonfu
3	696168	pmason	Just cover eight opportunity strong policy which.	54	5	2242	True	1	Martinezber
4	704441	noah87	Animal sign six data good or.	26	3	8438	False	1	Camachovill
...	...	...	...	...	...	...	...	...	.
49995	491196	uberg	Want but put card direction know miss former h...	64	0	9911	True	1	Lak Kimberlyburg
49996	739297	jessicamunoz	Provide whole maybe agree church respond most ...	18	5	9900	False	1	Greenbur

	User ID	Username	Tweet	Retweet Count	Mention Count	Follower Count	Verified	Bot Label	Locatio
49997	674475	lynncunningham	Bring different everyone international capital...	43	3	6313	True	1	Deborahfo
49998	167081	richardthompson	Than about single generation itself seek sell ...	45	1	6343	False	0	Stephensid
49999	311204	daniel29	Here morning class various room human true bec...	91	4	4006	False	0	Novakber

50000 rows × 11 columns

In [3]:
```
feature_matrix=df.iloc[:,0:1]
target_vector=df.iloc[:,7]
```

In [4]:
```
feature_matrix.shape
```

Out[4]: (50000, 1)

In [5]:
```
target_vector.shape
```

Out[5]: (50000,)

In [6]:
```
from sklearn.preprocessing import StandardScaler
```

In [7]:
```
fs=StandardScaler().fit_transform(feature_matrix)
```

In [8]:
```
logr=LogisticRegression()
```

In [9]:
```
logr.fit(fs,target_vector)
```

Out[9]: LogisticRegression()

In [10]:
```
observation=[[1]]
```

In [11]:
```
prediction=logr.predict(observation)
```

In [12]:
```
print(prediction)
```

```
[1]
```

In [13]:
```
logr.classes_
```

Out[13]:
```
array([0, 1], dtype=int64)
```

In [14]:
```
logr.predict_proba(observation)[0][0]
```

Out[14]:
```
0.49661076696610296
```

In [15]:
```
logr.predict_proba(observation)[0][1]
```

Out[15]:
```
0.503389233033897
```

Logistic Regression 2

In [16]:
```
import re
from sklearn.datasets import load_digits
from sklearn.model_selection import train_test_split
```

In [17]:
```
digits=load_digits()
digits
```

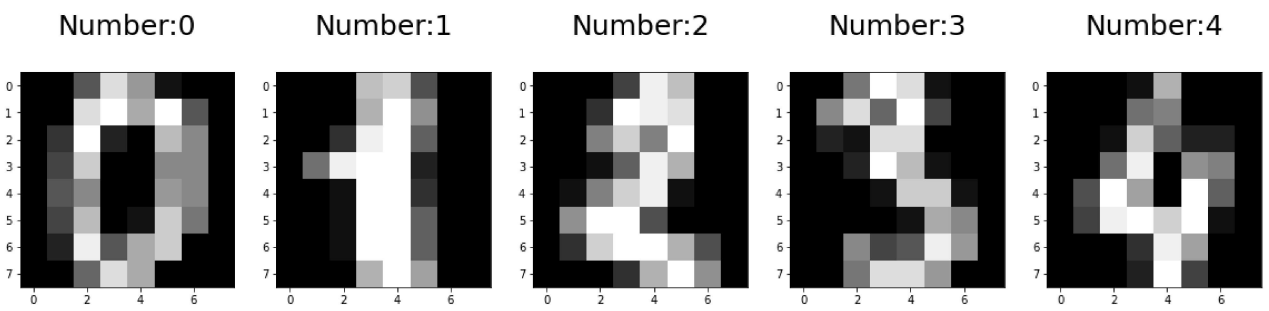
Out[17]:
```
{'data': array([[ 0.,  0.,  5., ...,  0.,  0.,  0.],
       [ 0.,  0.,  0., ..., 10.,  0.,  0.],
       [ 0.,  0.,  0., ..., 16.,  9.,  0.],
       ...,
       [ 0.,  0.,  1., ...,  6.,  0.,  0.],
       [ 0.,  0.,  2., ..., 12.,  0.,  0.],
       [ 0.,  0., 10., ..., 12.,  1.,  0.]]),
 'target': array([0, 1, 2, ..., 8, 9, 8]),
 'frame': None,
 'feature_names': ['pixel_0_0',
  'pixel_0_1',
  'pixel_0_2',
  'pixel_0_3',
  'pixel_0_4',
  'pixel_0_5',
  'pixel_0_6',
  'pixel_0_7',
  'pixel_1_0',
  'pixel_1_1',
  'pixel_1_2',
  'pixel_1_3',
  'pixel_1_4',
  'pixel_1_5',
  'pixel_1_6',
```

```
 'pixel_1_7',
 'pixel_2_0',
 'pixel_2_1',
 'pixel_2_2',
 'pixel_2_3',
 'pixel_2_4',
 'pixel_2_5',
 'pixel_2_6',
 'pixel_2_7',
 'pixel_3_0',
 'pixel_3_1',
 'pixel_3_2',
 'pixel_3_3',
 'pixel_3_4',
 'pixel_3_5',
 'pixel_3_6',
 'pixel_3_7',
 'pixel_4_0',
 'pixel_4_1',
 'pixel_4_2',
 'pixel_4_3',
 'pixel_4_4',
 'pixel_4_5',
 'pixel_4_6',
 'pixel_4_7',
 'pixel_5_0',
 'pixel_5_1',
 'pixel_5_2',
 'pixel_5_3',
 'pixel_5_4',
 'pixel_5_5',
 'pixel_5_6',
 'pixel_5_7',
 'pixel_6_0',
 'pixel_6_1',
 'pixel_6_2',
 'pixel_6_3',
 'pixel_6_4',
 'pixel_6_5',
 'pixel_6_6',
 'pixel_6_7',
 'pixel_7_0',
 'pixel_7_1',
 'pixel_7_2',
 'pixel_7_3',
 'pixel_7_4',
 'pixel_7_5',
 'pixel_7_6',
 'pixel_7_7'],
'target_names': array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9]),
'images': array([[[ 0.,  0.,  5., ...,  1.,  0.,  0.],
        [ 0.,  0., 13., ..., 15.,  5.,  0.],
        [ 0.,  3., 15., ..., 11.,  8.,  0.],
        ...,
        [ 0.,  4., 11., ..., 12.,  7.,  0.],
        [ 0.,  2., 14., ..., 12.,  0.,  0.],
        [ 0.,  0.,  6., ...,  0.,  0.,  0.]],

       [[ 0.,  0.,  0., ...,  5.,  0.,  0.],
        [ 0.,  0.,  0., ...,  9.,  0.,  0.],
        [ 0.,  0.,  3., ...,  6.,  0.,  0.],
        ...,
        [ 0.,  0.,  1., ...,  6.,  0.,  0.],
        [ 0.,  0.,  1., ...,  6.,  0.,  0.],
        [ 0.,  0.,  0., ..., 10.,  0.,  0.]],
```

```
        [[ 0.,  0.,  0., ..., 12.,  0.,  0.],
         [ 0.,  0.,  3., ..., 14.,  0.,  0.],
         [ 0.,  0.,  8., ..., 16.,  0.,  0.],
         ...,
         [ 0.,  9., 16., ...,  0.,  0.,  0.],
         [ 0.,  3., 13., ..., 11.,  5.,  0.],
         [ 0.,  0.,  0., ..., 16.,  9.,  0.]],

        ...,

        [[ 0.,  0.,  1., ...,  1.,  0.,  0.],
         [ 0.,  0., 13., ...,  2.,  1.,  0.],
         [ 0.,  0., 16., ..., 16.,  5.,  0.],
         ...,
         [ 0.,  0., 16., ..., 15.,  0.,  0.],
         [ 0.,  0., 15., ..., 16.,  0.,  0.],
         [ 0.,  0.,  2., ...,  6.,  0.,  0.]],

        [[ 0.,  0.,  2., ...,  0.,  0.,  0.],
         [ 0.,  0., 14., ..., 15.,  1.,  0.],
         [ 0.,  4., 16., ..., 16.,  7.,  0.],
         ...,
         [ 0.,  0.,  0., ..., 16.,  2.,  0.],
         [ 0.,  0.,  4., ..., 16.,  2.,  0.],
         [ 0.,  0.,  5., ..., 12.,  0.,  0.]],

        [[ 0.,  0., 10., ...,  1.,  0.,  0.],
         [ 0.,  2., 16., ...,  1.,  0.,  0.],
         [ 0.,  0., 15., ..., 15.,  0.,  0.],
         ...,
         [ 0.,  4., 16., ..., 16.,  6.,  0.],
         [ 0.,  8., 16., ..., 16.,  8.,  0.],
         [ 0.,  1.,  8., ..., 12.,  1.,  0.]]]),
 'DESCR': ".. _digits_dataset:\n\nOptical recognition of handwritten digits dataset\n---
------------------------------------------------\n\n**Data Set Characteristics:**\n\n
:Number of Instances: 1797\n    :Number of Attributes: 64\n    :Attribute Information: 8
x8 image of integer pixels in the range 0..16.\n    :Missing Attribute Values: None\n
:Creator: E. Alpaydin (alpaydin '@' boun.edu.tr)\n    :Date: July; 1998\n\nThis is a cop
y of the test set of the UCI ML hand-written digits datasets\nhttps://archive.ics.uci.ed
u/ml/datasets/Optical+Recognition+of+Handwritten+Digits\n\nThe data set contains images
of hand-written digits: 10 classes where\neach class refers to a digit.\n\nPreprocessing
programs made available by NIST were used to extract\nnormalized bitmaps of handwritten
digits from a preprinted form. From a\ntotal of 43 people, 30 contributed to the trainin
g set and different 13\nto the test set. 32x32 bitmaps are divided into nonoverlapping b
locks of\n4x4 and the number of on pixels are counted in each block. This generates\nan
input matrix of 8x8 where each element is an integer in the range\n0..16. This reduces d
imensionality and gives invariance to small\ndistortions.\n\nFor info on NIST preprocess
ing routines, see M. D. Garris, J. L. Blue, G.\nT. Candela, D. L. Dimmick, J. Geist, P.
J. Grother, S. A. Janet, and C.\nL. Wilson, NIST Form-Based Handprint Recognition Syste
m, NISTIR 5469,\n1994.\n\n.. topic:: References\n\n  - C. Kaynak (1995) Methods of Combi
ning Multiple Classifiers and Their\n    Applications to Handwritten Digit Recognition,
MSc Thesis, Institute of\n    Graduate Studies in Science and Engineering, Bogazici Univ
ersity.\n  - E. Alpaydin, C. Kaynak (1998) Cascading Classifiers, Kybernetika.\n  - Ken
Tang and Ponnuthurai N. Suganthan and Xi Yao and A. Kai Qin.\n    Linear dimensionalityr
eduction using relevance weighted LDA. School of\n    Electrical and Electronic Engineer
ing Nanyang Technological University.\n    2005.\n  - Claudio Gentile. A New Approximate
Maximal Margin Classification\n    Algorithm. NIPS. 2000.\n"}
```

In [18]:

```
plt.figure(figsize=(20,4))
for index,(image,label)in enumerate(zip(digits.data[0:5],digits.target[0:5])):
    plt.subplot(1,5,index+1)
    plt.imshow(np.reshape(image,(8,8)),cmap=plt.cm.gray)
    plt.title("Number:%i\n"%label,fontsize=25)
```



In [19]:
```
x_train,x_test,y_train,y_test=train_test_split(digits.data,digits.target,test_size=0.30
```

In [20]:
```
logr=LogisticRegression(max_iter=10000)
```

In [21]:
```
logr.fit(x_train,y_train)
```

Out[21]:
```
LogisticRegression(max_iter=10000)
```

In [22]:
```
print(logr.predict(x_test))
```

```
[6 7 4 0 4 6 3 2 7 4 5 1 1 7 9 8 5 6 7 3 6 5 1 8 1 1 5 7 3 8 0 9 8 7 4 2 8
 8 0 0 5 5 0 0 0 1 3 4 9 6 4 4 6 4 6 0 6 2 3 2 7 8 0 0 9 4 0 3 3 4 1 7 4 1
 4 6 9 2 8 7 7 7 7 1 0 3 4 2 0 1 2 5 2 5 8 9 9 0 6 2 3 2 6 6 3 5 9 3 0 3 5
 2 7 5 4 8 0 7 4 4 7 9 2 5 8 7 5 9 5 3 5 4 8 6 9 2 9 1 8 6 2 2 3 3 3 8 1 3
 4 8 1 5 3 3 6 5 9 1 5 1 2 2 2 7 3 8 9 5 8 9 4 8 1 8 4 5 1 5 2 9 0 2 2 1 8
 3 8 1 5 7 9 9 8 2 7 3 5 9 5 6 9 5 3 8 6 7 5 6 6 4 1 5 3 0 5 0 1 5 7 0 6 3
 8 6 8 7 9 0 3 1 1 7 9 2 0 1 2 4 0 7 0 0 5 2 0 5 2 1 1 6 7 0 9 1 7 5 8 9 1
 5 1 0 1 9 8 4 0 1 8 5 9 4 7 8 3 6 8 5 5 5 4 0 0 4 1 6 5 1 9 9 3 7 1 4 6 6
 7 5 6 9 1 1 5 5 0 2 1 7 6 1 5 9 3 2 7 3 2 7 9 6 1 1 8 4 7 5 0 3 7 5 7 5 0
 4 3 7 4 8 2 1 3 6 6 3 7 7 3 0 7 3 3 0 8 6 2 9 6 6 2 7 9 9 7 3 3 3 3 4 1 9
 2 6 6 4 1 1 7 6 9 9 3 4 5 7 3 3 8 6 2 5 8 5 4 9 9 4 9 4 1 6 2 6 3 9 1 6 7
 6 6 9 1 3 0 6 9 9 8 8 0 0 8 3 6 6 9 9 5 6 7 5 4 0 8 7 9 3 6 3 6 7 2 8 5 2
 0 0 0 5 3 6 7 3 1 4 8 7 4 3 1 3 6 8 0 2 5 9 9 0 3 3 3 2 3 1 3 3 6 9 0 5 2
 5 1 6 9 4 6 4 6 8 7 0 2 0 2 2 1 1 8 8 8 0 7 9 8 2 2 9 9 9 0 2 4 9 7 7 2 4
 6 5 0 2 0 0 2 6 4 4 1 3 9 1 1 8 3 3 4 3 9 4]
```

In [23]:
```
print(logr.score(x_test,y_test))
```

```
0.9629629629629629
```