In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [2]:
```
```

In [3]:
```
df = pd.read_csv("C3_bot_detection_data.csv")
```

Out[3]:

	User ID	Username	Tweet	Retweet Count	Mention Count	Follower Count	Verified	Bot Label	Lo
0	132131	flong	Station activity person against natural majori...	85	1	2353	False	1	Adł
1	289683	hinesstephanie	Authority research natural life material staff...	55	5	9617	True	0	Sanc
2	779715	roberttran	Manage whose quickly especially foot none to g...	6	2	4363	True	0	Harri
3	696168	pmason	Just cover eight opportunity strong policy which.	54	5	2242	True	1	Martin
4	704441	noah87	Animal sign six data good or.	26	3	8438	False	1	Camac
...	...	...	...	...	...	...	...	...	
49995	491196	uberg	Want but put card direction know miss former h...	64	0	9911	True	1	Kimberl
49996	739297	jessicamunoz	Provide whole maybe agree church respond most ...	18	5	9900	False	1	Gre
49997	674475	lynncunningham	Bring different everyone international capital...	43	3	6313	True	1	Debo
49998	167081	richardthompson	Than about single generation itself seek sell ...	45	1	6343	False	0	Steph

	User ID	Username	Tweet	Retweet Count	Mention Count	Follower Count	Verified	Bot Label	Lo
			Here morning class						

In [4]:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50000 entries, 0 to 49999
Data columns (total 11 columns):
 #   Column          Non-Null Count  Dtype 
---  ------          --------------  ----- 
 0   User ID         50000 non-null  int64 
 1   Username        50000 non-null  object
 2   Tweet           50000 non-null  object
 3   Retweet Count   50000 non-null  int64 
 4   Mention Count   50000 non-null  int64 
 5   Follower Count  50000 non-null  int64 
 6   Verified        50000 non-null  bool  
 7   Bot Label       50000 non-null  int64 
 8   Location        50000 non-null  object
 9   Created At      50000 non-null  object
 10  Hashtags        41659 non-null  object
dtypes: bool(1), int64(5), object(5)
memory usage: 3.9+ MB
```

In [5]:

Out[5]:

```
Index(['User ID', 'Username', 'Tweet', 'Retweet Count', 'Mention Count',
       'Follower Count', 'Verified', 'Bot Label', 'Location', 'Created At',
       'Hashtags'],
      dtype='object')
```

In [6]:

```
f_m=df[['User ID','Retweet Count', 'Mention Count',
       'Follower Count','Bot Label']]
```

In [7]:

Out[7]:

```
(50000, 5)
```

In [8]:

Out[8]:

```
(50000,)
```

In [9]:

In [10]:

In [11]:

```
logr=LogisticRegression()
```

Out[11]:

```
LogisticRegression()
```

In [12]:

In [13]:
```
prediction=logr.predict(observation)
```

Out[13]:
```
array([ True])
```

In [14]:

Out[14]:
```
array([False,  True])
```

In [15]:

Out[15]:
```
0.4875957520146553
```

In [16]:

Out[16]:
```
0.5124042479853447
```

RANDOM FOREST

In [17]:

Out[17]:
```
True     25004
False    24996
Name: Verified, dtype: int64
```

In [18]:
```
x=df[['User ID','Retweet Count', 'Mention Count',
       'Follower Count','Bot Label']]
y=df['Verified']
```

In [19]:
```
g1={"Verified":{'True':1,'False':2}}
df=df.replace(g1)
```

Out[19]:

	User ID	Username	Tweet	Retweet Count	Mention Count	Follower Count	Verified	Bot Label	Lo
0	132131	flong	Station activity person against natural majori...	85	1	2353	False	1	Adł
1	289683	hinesstephanie	Authority research natural life material staff...	55	5	9617	True	0	Sand
2	779715	roberttran	Manage whose quickly especially foot none to g...	6	2	4363	True	0	Harri
3	696168	pmason	Just cover eight opportunity strong policy which.	54	5	2242	True	1	Martin
4	704441	noah87	Animal sign six data good or.	26	3	8438	False	1	Camac
...	...	...	...	...	...	...	...	...	
49995	491196	uberg	Want but put card direction know miss former h...	64	0	9911	True	1	Kimberl
49996	739297	jessicamunoz	Provide whole maybe agree church respond most ...	18	5	9900	False	1	Gre
49997	674475	lynncunningham	Bring different everyone international capital...	43	3	6313	True	1	Debo
49998	167081	richardthompson	Than about single generation itself seek sell ...	45	1	6343	False	0	Steph

	User ID	Username	Tweet	Retweet Count	Mention Count	Follower Count	Verified	Bot Label	Lo
49999	311204	daniel29	Here morning class various	91	4	4006	False	0	Nov

In [20]:
```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,train_size=0.70)
```

In [21]:
```
rfc=RandomForestClassifier()
```

Out[21]:
```
RandomForestClassifier()
```

In [22]:
```
parameters={'max_depth':[1,2,3,4,5],
            'min_samples_leaf':[5,10,15,20,25],
```

In [23]:
```
from sklearn.model_selection import GridSearchCV
grid_search =GridSearchCV(estimator=rfc,param_grid=parameters,cv=2,scoring="ac
```

Out[23]:
```
GridSearchCV(cv=2, estimator=RandomForestClassifier(),
             param_grid={'max_depth': [1, 2, 3, 4, 5],
                         'min_samples_leaf': [5, 10, 15, 20, 25],
                         'n_estimators': [10, 20, 30, 40, 50]},
             scoring='accuracy')
```

In [24]:

Out[24]:
```
0.5050571428571429
```

In [25]:

In [26]:
```
from sklearn.tree import plot_tree
plt.figure(figsize=(80,40))
plot_tree(rfc_best.estimators_[5],feature_names=x.columns,class_names=['Yes','
```

Out[26]:
```
[Text(2232.0, 1902.6000000000001, 'User ID <= 144062.5\ngini = 0.5\nsamples =
22121\nvalue = [17478, 17522]\nclass = No'),
 Text(1116.0, 1359.0, 'Mention Count <= 2.5\ngini = 0.498\nsamples = 1047\nva
lue = [857, 753]\nclass = Yes'),
 Text(558.0, 815.4000000000001, 'Retweet Count <= 79.5\ngini = 0.5\nsamples =
517\nvalue = [394, 373]\nclass = Yes'),
 Text(279.0, 271.79999999999995, 'gini = 0.498\nsamples = 432\nvalue = [343,
307]\nclass = Yes'),
 Text(837.0, 271.79999999999995, 'gini = 0.492\nsamples = 85\nvalue = [51, 6
6]\nclass = No'),
 Text(1674.0, 815.4000000000001, 'User ID <= 139775.5\ngini = 0.495\nsamples
= 530\nvalue = [463, 380]\nclass = Yes'),
 Text(1395.0, 271.79999999999995, 'gini = 0.492\nsamples = 473\nvalue = [427,
332]\nclass = Yes'),
 Text(1953.0, 271.79999999999995, 'gini = 0.49\nsamples = 57\nvalue = [36, 4
8]\nclass = No'),
 Text(3348.0, 1359.0, 'Mention Count <= 2.5\ngini = 0.5\nsamples = 21074\nval
ue = [16621, 16769]\nclass = No'),
 Text(2790.0, 815.4000000000001, 'Follower Count <= 553.5\ngini = 0.5\nsample
s = 10540\nvalue = [8270, 8474]\nclass = No'),
 Text(2511.0, 271.79999999999995, 'gini = 0.49\nsamples = 582\nvalue = [408,
540]\nclass = No'),
 Text(3069.0, 271.79999999999995, 'gini = 0.5\nsamples = 9958\nvalue = [7862,
7934]\nclass = No'),
 Text(3906.0, 815.4000000000001, 'Follower Count <= 2044.5\ngini = 0.5\nsampl
es = 10534\nvalue = [8351, 8295]\nclass = Yes'),
 Text(3627.0, 271.79999999999995, 'gini = 0.499\nsamples = 2147\nvalue = [178
0, 1596]\nclass = Yes'),
 Text(4185.0, 271.79999999999995, 'gini = 0.5\nsamples = 8387\nvalue = [6571,
6699]\nclass = No')]
```