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

```
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	Location
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	Harrisonfur
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	Location
49997	674475	lynncunningham	Bring different everyone international capital...	43	3	6313	True	1	Deborahfor
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	Novakberi

50000 rows × 11 columns

In [3]:
```
df.columns
```

Out[3]: Index(['User ID', 'Username', 'Tweet', 'Retweet Count', 'Mention Count',
       'Follower Count', 'Verified', 'Bot Label', 'Location', 'Created At',
       'Hashtags'],
      dtype='object')

In [4]:
```
df.info()
```

```
<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]:
```
df['Verified'].value_counts()
```

Out[5]: True   25004
        False   24996
        Name: Verified, dtype: int64

In [6]:
```python
df['Verified'].value_counts()
```

Out[6]:
```
True      25004
False     24996
Name: Verified, dtype: int64
```

In [7]:
```python
x=df[['User ID',  'Retweet Count', 'Mention Count',
      'Follower Count', 'Bot Label']]
y=df['Verified']
```

In [8]:
```python
g1={"Verified":{'False':1,'True':2}}
df=df.replace(g1)
print(df)
```

```
        User ID          Username  \
0        132131             flong
1        289683   hinesstephanie
2        779715       roberttran
3        696168           pmason
4        704441           noah87
...         ...              ...
49995    491196            uberg
49996    739297      jessicamunoz
49997    674475   lynncunningham
49998    167081   richardthompson
49999    311204         daniel29

                                               Tweet  Retweet Count  \
0      Station activity person against natural majori...             85
1      Authority research natural life material staff...             55
2      Manage whose quickly especially foot none to g...              6
3      Just cover eight opportunity strong policy which.             54
4                        Animal sign six data good or.             26
...                                              ...            ...
49995  Want but put card direction know miss former h...             64
49996  Provide whole maybe agree church respond most ...             18
49997  Bring different everyone international capital...             43
49998  Than about single generation itself seek sell ...             45
49999  Here morning class various room human true bec...             91

       Mention Count  Follower Count  Verified  Bot Label           Location  \
0                  1            2353     False          1          Adkinston
1                  5            9617      True          0          Sanderston
2                  2            4363      True          0        Harrisonfurt
3                  5            2242      True          1        Martinezberg
4                  3            8438     False          1        Camachoville
...              ...             ...       ...        ...                ...
49995              0            9911      True          1   Lake Kimberlyburgh
49996              5            9900     False          1          Greenbury
49997              3            6313      True          1        Deborahfort
49998              1            6343     False          0        Stephenside
49999              4            4006     False          0          Novakberg

              Created At               Hashtags
0      2020-05-11 15:29:50                    NaN
1      2022-11-26 05:18:10              both live
2      2022-08-08 03:16:54            phone ahead
3      2021-08-14 22:27:05         ever quickly new I
4      2020-04-13 21:24:21         foreign mention
...                    ...                    ...
```

```
49995   2023-04-20 11:06:26   teach quality ten education any
49996   2022-10-18 03:57:35          add walk among believe
49997   2020-07-08 03:54:08        onto admit artist first
49998   2022-03-22 12:13:44                             star
49999   2022-12-03 06:11:07                             home

[50000 rows x 11 columns]
```

In [9]:
```python
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.70)
```

In [10]:
```python
from sklearn.ensemble import RandomForestClassifier
rfc=RandomForestClassifier()
rfc.fit(x_train,y_train)
```

Out[10]: RandomForestClassifier()

In [11]:
```python
parameters= {
    "max_depth":[1,2,3,4,5],
    "min_samples_leaf":[5,10,15,20,25],
    'n_estimators':[10,20,30,40,50]
}
```

In [12]:
```python
from sklearn.model_selection import GridSearchCV
grid_search=GridSearchCV(estimator=rfc,param_grid=parameters,cv=2,scoring="accuracy")
grid_search.fit(x_train,y_train)
```

Out[12]:
```
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 [13]:
```python
grid_search.best_score_
```

Out[13]: 0.5065333333333333

In [14]:
```python
rfc_best=grid_search.best_estimator_
```

In [15]:
```python
from sklearn.tree import plot_tree
plt.figure(figsize=(80,40))
plot_tree(rfc_best.estimators_[5],feature_names=x.columns,class_names=['Yes','No'],fill
```
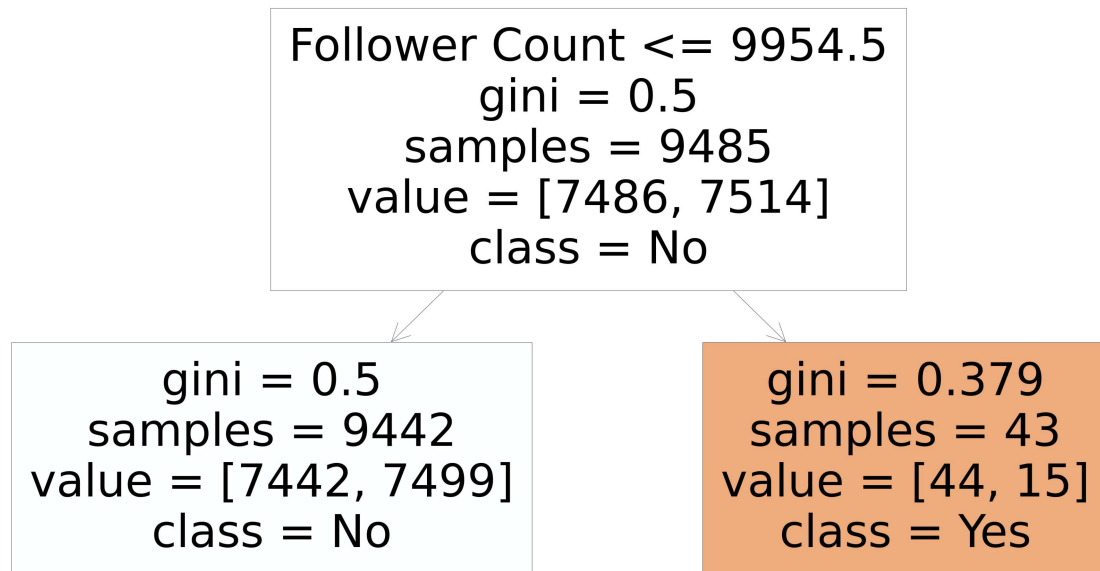
Out[15]:
```
[Text(2232.0, 1630.8000000000002, 'Follower Count <= 9954.5\ngini = 0.5\nsamples = 9485
\nvalue = [7486, 7514]\nclass = No'),
 Text(1116.0, 543.5999999999999, 'gini = 0.5\nsamples = 9442\nvalue = [7442, 7499]\nclas
s = No'),
 Text(3348.0, 543.5999999999999, 'gini = 0.379\nsamples = 43\nvalue = [44, 15]\nclass =
Yes')]
```

Follower Count <= 9954.5
gini = 0.5
samples = 9485
value = [7486, 7514]
class = No

gini = 0.5
samples = 9442
value = [7442, 7499]
class = No

gini = 0.379
samples = 43
value = [44, 15]
class = Yes

In [ ]: