In [1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:

```
df=pd.read_csv('bot.csv')
df
```

Out[2]:

| | User ID | Username | Tweet | Retweet Count | Mention Count | Follower Count | Verified | Bot Label | Location | Cre |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 132131 | flong | Station activity person against natural majori... | 85 | 1 | 2353 | False | 1 | Adkinston | 15: |
| 1 | 289683 | hinesstephanie | Authority research natural life material staff... | 55 | 5 | 9617 | True | 0 | Sanderston | 05: |
| 2 | 779715 | roberttran | Manage whose quickly especially | 6 | 2 | 4363 | True | 0 | Harrisonfurt | |

In [19]:

```
df.columns
```

Out[19]:

```
Index(['User ID', 'Username', 'Tweet', 'Retweet Count', 'Mention Count',
       'Follower Count', 'Verified', 'Bot Label', 'Location', 'Created A
t',
       'Hashtags'],
      dtype='object')
```

In [20]:

```
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 [37]:

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

Out[37]:

```
True     25004
False    24996
Name: Verified, dtype: int64
```

In [43]:

```
x=df[['User ID', 'Retweet Count', 'Mention Count',
       'Follower Count']]
y=df['Verified']
```

In [44]:

```
d={"Verified":{'True':1,'False ':2}}
df=df.replace(df)
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...             28
1      Authority research natural life material staff...             23
2      Manage whose quickly especially foot none to g...             57
3      Just cover eight opportunity strong policy which.             43
4                          Animal sign six data good or.             28
...                                                  ...            ...
49995  Want but put card direction know miss former h...             37
49996  Provide whole maybe agree church respond most ...             85
49997  Bring different everyone international capital...              8
49998  Than about single generation itself seek sell ...             85
49999  Here morning class various room human true bec...              3

       Mention Count  Follower Count  Verified  Bot Label            Locat
ion  \
0                  3            2937     False          0           Adkins
ton
1                  5            3512      True          1          Sanders
ton
2                  2            7465      True          1        Harrisonf
urt
3                  5            5906      True          0        Martinezb
erg
4                  3            3139     False          0        Camachovi
lle
...              ...             ...       ...        ...
...
49995              4            4962      True          0  Lake Kimberlybu
rgh
49996              5            2014     False          0           Greenb
ury
49997              3            2642      True          0         Deborahf
ort
49998              3            6812     False          1         Stephens
ide
49999              4            5119     False          1           Novakb
erg

                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 [45]:

```
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 [46]:

```
from sklearn.ensemble import RandomForestClassifier
rfc=RandomForestClassifier()
rfc.fit(x_train,y_train)
```

Out[46]:

```
RandomForestClassifier()
```

# Depth of Tree

In [47]:

```
parameters={"max_depth":[1,2,3,4,5],"min_samples_leaf":[5,23,45,76,78],'n_estimators':[10
```

# Cross Validate

In [48]:

```
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[48]:

```
GridSearchCV(cv=2, estimator=RandomForestClassifier(),
             param_grid={'max_depth': [1, 2, 3, 4, 5],
                         'min_samples_leaf': [5, 23, 45, 76, 78],
                         'n_estimators': [10, 23, 45, 65, 7]},
             scoring='accuracy')
```

# Score

In [49]:

```
grid_search.best_score_
```

Out[49]:

```
0.5092
```

In [50]:

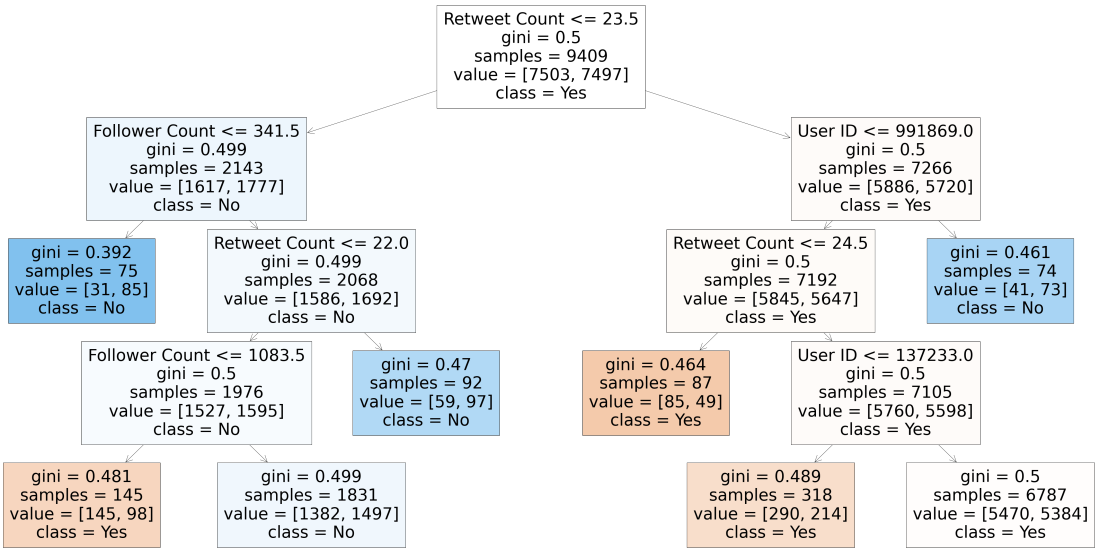
```
rfc_best=grid_search.best_estimator_
```

In [51]:

```
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'],filled
```

Out[51]:

```
[Text(2232.0, 1956.96, 'Retweet Count <= 23.5\ngini = 0.5\nsamples = 9409
\nvalue = [7503, 7497]\nclass = Yes'),
 Text(892.8, 1522.0800000000002, 'Follower Count <= 341.5\ngini = 0.499\ns
amples = 2143\nvalue = [1617, 1777]\nclass = No'),
 Text(446.4, 1087.2, 'gini = 0.392\nsamples = 75\nvalue = [31, 85]\nclass
= No'),
 Text(1339.1999999999998, 1087.2, 'Retweet Count <= 22.0\ngini = 0.499\nsa
mples = 2068\nvalue = [1586, 1692]\nclass = No'),
 Text(892.8, 652.3200000000002, 'Follower Count <= 1083.5\ngini = 0.5\nsam
ples = 1976\nvalue = [1527, 1595]\nclass = No'),
 Text(446.4, 217.44000000000005, 'gini = 0.481\nsamples = 145\nvalue = [14
5, 98]\nclass = Yes'),
 Text(1339.1999999999998, 217.44000000000005, 'gini = 0.499\nsamples = 183
1\nvalue = [1382, 1497]\nclass = No'),
 Text(1785.6, 652.3200000000002, 'gini = 0.47\nsamples = 92\nvalue = [59,
97]\nclass = No'),
 Text(3571.2, 1522.0800000000002, 'User ID <= 991869.0\ngini = 0.5\nsample
s = 7266\nvalue = [5886, 5720]\nclass = Yes'),
 Text(3124.7999999999997, 1087.2, 'Retweet Count <= 24.5\ngini = 0.5\nsamp
les = 7192\nvalue = [5845, 5647]\nclass = Yes'),
 Text(2678.3999999999996, 652.3200000000002, 'gini = 0.464\nsamples = 87\n
value = [85, 49]\nclass = Yes'),
 Text(3571.2, 652.3200000000002, 'User ID <= 137233.0\ngini = 0.5\nsamples
= 7105\nvalue = [5760, 5598]\nclass = Yes'),
 Text(3124.7999999999997, 217.44000000000005, 'gini = 0.489\nsamples = 318
\nvalue = [290, 214]\nclass = Yes'),
 Text(4017.6, 217.44000000000005, 'gini = 0.5\nsamples = 6787\nvalue = [54
70, 5384]\nclass = Yes'),
 Text(4017.6, 1087.2, 'gini = 0.461\nsamples = 74\nvalue = [41, 73]\nclass
= No')]
```



In [ ]:

```
```