

```
In [81]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
.
```

```
In [119]: df=pd.read_csv(r"C:\USERS\user\Downloads\C3_bot_detection_data - C3_bot_detect
```

```
Out[119]:
```

	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	Adl
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
		Here morning class						

In [120]:

Out[120]: Index(['User ID', 'Username', 'Tweet', 'Retweet Count', 'Mention Count', 'Follower Count', 'Verified', 'Bot Label', 'Location', 'Created At', 'Hashtags'], dtype='object')

```
In [121]: df=df.head(10)
```

```
Out[121]:
```

	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	Adkinstr
1	289683	hinesstephanie	Authority research natural life material staff...	55	5	9617	True	0	Sanderstr
2	779715	roberttran	Manage whose quickly especially foot none to g...	6	2	4363	True	0	Harrisonft
3	696168	pmason	Just cover eight opportunity strong policy which.	54	5	2242	True	1	Martinezbe
4	704441	noah87	Animal sign six data good or.	26	3	8438	False	1	Camachovil
5	570928	james00	See wonder travel this suffer less yard office...	41	4	3792	True	1	West Cheyenr
6	734182	leonard00	Door final sound my guess building rich.	54	0	10	True	0	South Dona
7	107312	lesterdaniel	Job phone price magazine worry stage check view.	64	0	1442	False	1	Smithhav
8	549888	kimberlymorris	Eye rest prove mission show floor.	25	2	836	False	0	Lal Brittanyvil
9	117640	schmittjill	Add letter year performance western what cultu...	67	3	6523	False	1	We Hannahborou

In [122]: `a=df[['User ID','Retweet Count','Mention Count','Follower Count','Bot Label'],'`

Out[122]:

	User ID	Retweet Count	Mention Count	Follower Count	Bot Label	Verified
0	132131	85	1	2353	1	False
1	289683	55	5	9617	0	True
2	779715	6	2	4363	0	True
3	696168	54	5	2242	1	True
4	704441	26	3	8438	1	False
5	570928	41	4	3792	1	True
6	734182	54	0	10	0	True
7	107312	64	0	1442	1	False
8	549888	25	2	836	0	False
9	117640	67	3	6523	1	False

In [123]:

Out[123]:

```
False    5
True     5
Name: Verified, dtype: int64
```

In [124]: `x=a.drop('Verified',axis=1)`

In [125]: `g1={"Verified":{"False":1,'True':2}}`
`a=a.replace(g1)`

	User ID	Retweet Count	Mention Count	Follower Count	Bot Label	Verified
0	132131	85	1	2353	1	False
1	289683	55	5	9617	0	True
2	779715	6	2	4363	0	True
3	696168	54	5	2242	1	True
4	704441	26	3	8438	1	False
5	570928	41	4	3792	1	True
6	734182	54	0	10	0	True
7	107312	64	0	1442	1	False
8	549888	25	2	836	0	False
9	117640	67	3	6523	1	False

In [126]: `from sklearn.model_selection import train_test_split`

In [127]: `from sklearn.ensemble import RandomForestClassifier`

```
rfc=RandomForestClassifier()
```

Out[127]: `RandomForestClassifier()`

In [128]: `parameters={'max_depth':[1,2,3,4,5],`
`'min_samples_leaf':[5,10,15,20,25],`

```
In [129]: from sklearn.model_selection import GridSearchCV
```

```
grid_search=GridSearchCV(estimator=rfc,param_grid=parameters,cv=2,scoring="acc
```

```
Out[129]: 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 [130]:
```

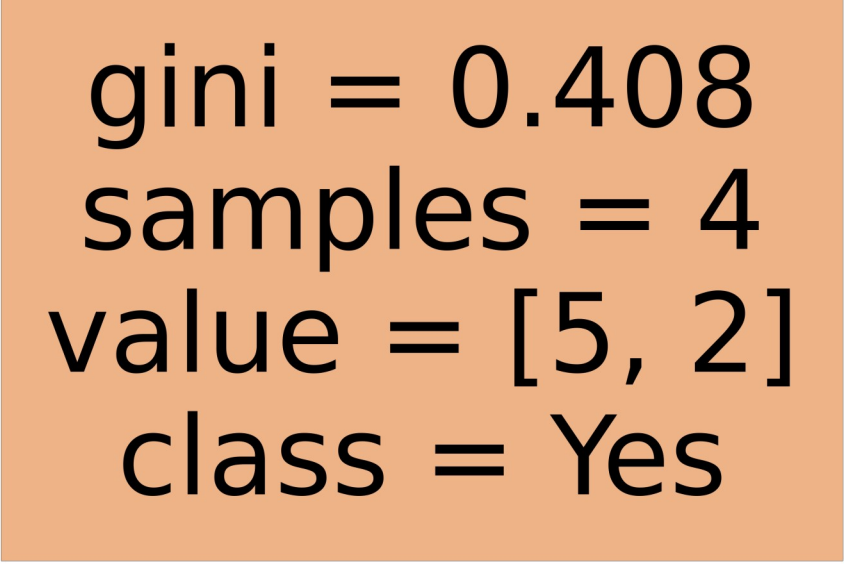
```
Out[130]: 0.5833333333333333
```

```
In [131]:
```

```
In [132]: from sklearn.tree import plot_tree
```

```
plt.figure(figsize=(80,40))
```

```
Out[132]: [Text(2232.0, 1087.2, 'gini = 0.408\nsamples = 4\nvalue = [5, 2]\nnclass = Yes
')]
```



gini = 0.408
samples = 4
value = [5, 2]
class = Yes

```
In [ ]:
```