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

In [160]:
```
data=pd.read_csv(r"C:\Users\user\Desktop\vicky\C2_train.gender_submission (1).csv"
```

In [161]:
```
data.fillna(value=1)
```

Out[161]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	1
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	1
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	1
...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	1
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	1.0	1	2	W./C. 6607	23.4500	1
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	1

891 rows × 12 columns

In [162]:
```
data.head()
```

Out[162]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN

In [163]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   PassengerId  891 non-null    int64  
 1   Survived     891 non-null    int64  
 2   Pclass       891 non-null    int64  
 3   Name         891 non-null    object 
 4   Sex          891 non-null    object 
 5   Age          714 non-null    float64
 6   SibSp        891 non-null    int64  
 7   Parch        891 non-null    int64  
 8   Ticket       891 non-null    object 
 9   Fare         891 non-null    float64
 10  Cabin        204 non-null    object 
 11  Embarked     889 non-null    object 
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB
```

In [164]:
```
data1=data[['PassengerId','Pclass','Sex','SibSp','Parch']]
```

In [165]:
```
data1['Sex'].value_counts()
```

Out[165]:
```
male      577
female    314
Name: Sex, dtype: int64
```

In [166]:
```
x=data1.drop('Sex',axis=1)
y=data1['Sex']
```

In [ ]:
```

```

In [167]:
```
g1={"Sex":{'male':1,'female':0,}}
data1=data1.replace(g1)
print(data1)
```

```
     PassengerId  Pclass  Sex  SibSp  Parch
0              1       3    1      1      0
1              2       1    0      1      0
2              3       3    0      0      0
3              4       1    0      1      0
4              5       3    1      0      0
..           ...     ...  ...    ...    ...
886          887       2    1      0      0
887          888       1    0      0      0
888          889       3    0      1      2
889          890       1    1      0      0
890          891       3    1      0      0

[891 rows x 5 columns]
```

In [168]:
```
from sklearn.model_selection import train_test_split
```

In [169]:
```
x_train,x_test,y_train,y_test=train_test_split(x,y,train_size=0.70)
```

In [170]:
```
from sklearn.ensemble import RandomForestClassifier
```

In [171]:
```
rfc=RandomForestClassifier()
rfc.fit(x_train,y_train)
```

Out[171]:
```
RandomForestClassifier()
```

In [172]:
```
parameters = {'max_depth':[1,2,3,4,5],
              'min_samples_leaf':[5,10,15,20,25],
              'n_estimators':[10,20,30,40,50]

}
```

In [173]:
```
from sklearn.model_selection import GridSearchCV
grid_search=GridSearchCV(estimator=rfc,param_grid=parameters,cv=2,scoring="accurac
grid_search.fit(x_train,y_train)
```

Out[173]:
```
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 [174]:
```
grid_search.best_score_
```

Out[174]:
```
0.6870001648940556
```

In [175]:
```
from sklearn.tree import plot_tree
```

In [176]:
```
rfc_best=grid_search.best_estimator_
```

In [177]:
```
plt.figure(figsize=(80,40))
plot_tree(rfc_best.estimators_[5],feature_names=x.columns,class_names=['Yes','No'],
```

Out[177]:
```
[Text(2232.0, 1956.96, 'SibSp <= 0.5\ngini = 0.451\nsamples = 398\nvalue = [214,
409]\nclass = No'),
 Text(1116.0, 1522.0800000000002, 'Parch <= 0.5\ngini = 0.391\nsamples = 266\nval
ue = [110, 302]\nclass = No'),
 Text(558.0, 1087.2, 'PassengerId <= 408.5\ngini = 0.342\nsamples = 238\nvalue =
[82, 292]\nclass = No'),
 Text(279.0, 652.3200000000002, 'Pclass <= 2.5\ngini = 0.412\nsamples = 112\nvalu
e = [52, 127]\nclass = No'),
 Text(139.5, 217.44000000000005, 'gini = 0.453\nsamples = 45\nvalue = [25, 47]\nc
lass = No'),
 Text(418.5, 217.44000000000005, 'gini = 0.377\nsamples = 67\nvalue = [27, 80]\nc
lass = No'),
 Text(837.0, 652.3200000000002, 'Pclass <= 1.5\ngini = 0.26\nsamples = 126\nvalue
= [30, 165]\nclass = No'),
 Text(697.5, 217.44000000000005, 'gini = 0.326\nsamples = 27\nvalue = [8, 31]\ncl
ass = No'),
 Text(976.5, 217.44000000000005, 'gini = 0.242\nsamples = 99\nvalue = [22, 134]\n
class = No'),
 Text(1674.0, 1087.2, 'Parch <= 1.5\ngini = 0.388\nsamples = 28\nvalue = [28, 10]
\nclass = Yes'),
 Text(1395.0, 652.3200000000002, 'Pclass <= 1.5\ngini = 0.463\nsamples = 15\nvalu
e = [14, 8]\nclass = Yes'),
 Text(1255.5, 217.44000000000005, 'gini = 0.496\nsamples = 6\nvalue = [6, 5]\ncla
ss = Yes'),
 Text(1534.5, 217.44000000000005, 'gini = 0.397\nsamples = 9\nvalue = [8, 3]\ncla
ss = Yes'),
 Text(1953.0, 652.3200000000002, 'PassengerId <= 283.5\ngini = 0.219\nsamples = 1
3\nvalue = [14, 2]\nclass = Yes'),
 Text(1813.5, 217.44000000000005, 'gini = 0.408\nsamples = 5\nvalue = [5, 2]\ncla
ss = Yes'),
 Text(2092.5, 217.44000000000005, 'gini = 0.0\nsamples = 8\nvalue = [9, 0]\nclass
= Yes'),
 Text(3348.0, 1522.0800000000002, 'Parch <= 1.5\ngini = 0.5\nsamples = 132\nvalue
= [104, 107]\nclass = No'),
 Text(2790.0, 1087.2, 'PassengerId <= 761.0\ngini = 0.493\nsamples = 106\nvalue =
[76, 97]\nclass = No'),
 Text(2511.0, 652.3200000000002, 'Parch <= 0.5\ngini = 0.472\nsamples = 94\nvalue
= [59, 96]\nclass = No'),
 Text(2371.5, 217.44000000000005, 'gini = 0.494\nsamples = 63\nvalue = [42, 52]\n
class = No'),
 Text(2650.5, 217.44000000000005, 'gini = 0.402\nsamples = 31\nvalue = [17, 44]\n
class = No'),
 Text(3069.0, 652.3200000000002, 'Pclass <= 1.5\ngini = 0.105\nsamples = 12\nvalu
e = [17, 1]\nclass = Yes'),
 Text(2929.5, 217.44000000000005, 'gini = 0.0\nsamples = 6\nvalue = [10, 0]\nclas
s = Yes'),
 Text(3208.5, 217.44000000000005, 'gini = 0.219\nsamples = 6\nvalue = [7, 1]\ncla
ss = Yes'),
 Text(3906.0, 1087.2, 'PassengerId <= 309.0\ngini = 0.388\nsamples = 26\nvalue =
[28, 10]\nclass = Yes'),
 Text(3627.0, 652.3200000000002, 'Pclass <= 2.5\ngini = 0.5\nsamples = 12\nvalue
= [8, 8]\nclass = Yes'),
 Text(3487.5, 217.44000000000005, 'gini = 0.49\nsamples = 5\nvalue = [3, 4]\nclas
s = No'),
 Text(3766.5, 217.44000000000005, 'gini = 0.494\nsamples = 7\nvalue = [5, 4]\ncla
ss = Yes'),
 Text(4185.0, 652.3200000000002, 'PassengerId <= 438.5\ngini = 0.165\nsamples = 1
4\nvalue = [20, 2]\nclass = Yes'),
```

```
 Text(4045.5, 217.44000000000005, 'gini = 0.0\nsamples = 5\nvalue = [9, 0]\nclass
= Yes'),
 Text(4324.5, 217.44000000000005, 'gini = 0.26\nsamples = 9\nvalue = [11, 2]\ncla
ss = Yes')]
```

In [ ]:

In [ ]: