

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('titanic.csv')
df
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
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	

In [3]:

```
df.columns
```

Out[3]:

```
Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp',
       'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked'],
      dtype='object')
```

In [4]:

```
df.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 [18]:

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

Out[18]:

```
male      577
female    314
Name: Sex, dtype: int64
```

In [19]:

```
x=df[['PassengerId', 'Survived', 'Pclass','Parch']]
y=df['Sex']
```

In [20]:

```
d={"Sex":{"male":1,'female':2}}
df=df.replace(df)
print(df)
```

	PassengerId	Survived	Pclass	\	Name	Sex	Age	SibSp
0	4	0	1					
1	5	1	1					
2	6	1	1					
3	7	1	1					
4	8	0	1					
..	...	...	...					
886	890	0	1					
887	891	1	1					
888	891	0	1					
889	891	1	1					
890	891	0	1					

  

		Name	Sex	Age	SibSp
0		Braund, Mr. Owen Harris	male	NaN	1
1	Cumings, Mrs. John Bradley (Florence Briggs Th...		female	NaN	1
2	Heikkinen, Miss. Laina		female	NaN	1
3	Futrelle, Mrs. Jacques Heath (Lily May Peel)		female	NaN	1
4	Allen, Mr. William Henry		male	NaN	1
..	...	...	...	...	...
886	Montvila, Rev. Juozas		male	NaN	1
887	Graham, Miss. Margaret Edith		female	NaN	1
888	Johnston, Miss. Catherine Helen "Carrie"		female	NaN	1
889	Behr, Mr. Karl Howell		male	NaN	1
890	Dooley, Mr. Patrick		male	NaN	1

  

	Parch	Ticket	Fare	Cabin	Embarked
0	0	A/5 21171	7.2500	NaN	S
1	0	PC 17599	71.2833	C85	C
2	0	STON/O2. 3101282	7.9250	NaN	S
3	0	113803	53.1000	C123	S
4	0	373450	8.0500	NaN	S
..	...	...	...	...	...
886	0	211536	31.2750	NaN	S
887	0	112053	27.7208	B42	S
888	0	W./C. 6607	23.4500	NaN	S
889	0	111369	27.7208	C148	C
890	0	370376	7.7500	NaN	Q

[891 rows x 12 columns]

In [21]:

```
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 [22]:

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

Out[22]:

```
RandomForestClassifier()
```

## Depth of Tree

In [23]:

```
parameters={"max_depth":[1,2,3,4,5],"min_samples_leaf":[5,23,45,76,78],'n_estimators':[10,20,30,40,50]}
```

## Cross Validate

In [24]:

```
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[24]:

```
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 [25]:

```
grid_search.best_score_
```

Out[25]:

```
0.7603523734709909
```

In [26]:

```
rfc_best=grid_search.best_estimator_
```

In [27]:

```
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[27]:

```
[Text(2697.0, 1956.96, 'PassengerId <= 655.5\ngini = 0.439\nsamples = 177\nvalue = [87, 180]\nclass = No'),
 Text(1488.0, 1522.0800000000002, 'Pclass <= 2.0\ngini = 0.465\nsamples = 129\nvalue = [70, 120]\nclass = No'),
 Text(744.0, 1087.2, 'Survived <= 0.5\ngini = 0.429\nsamples = 105\nvalue = [48, 106]\nclass = No'),
 Text(372.0, 652.3200000000002, 'PassengerId <= 552.0\ngini = 0.32\nsamples = 69\nvalue = [21, 84]\nclass = No'),
 Text(186.0, 217.44000000000005, 'gini = 0.282\nsamples = 61\nvalue = [16, 78]\nclass = No'),
 Text(558.0, 217.44000000000005, 'gini = 0.496\nsamples = 8\nvalue = [5, 6]\nclass = No'),
 Text(1116.0, 652.3200000000002, 'PassengerId <= 162.5\ngini = 0.495\nsamples = 36\nvalue = [27, 22]\nclass = Yes'),
 Text(930.0, 217.44000000000005, 'gini = 0.198\nsamples = 6\nvalue = [8, 1]\nclass = Yes'),
 Text(1302.0, 217.44000000000005, 'gini = 0.499\nsamples = 30\nvalue = [19, 21]\nclass = No'),
 Text(2232.0, 1087.2, 'Survived <= 0.5\ngini = 0.475\nsamples = 24\nvalue = [22, 14]\nclass = Yes'),
 Text(1860.0, 652.3200000000002, 'PassengerId <= 237.0\ngini = 0.231\nsamples = 11\nvalue = [2, 13]\nclass = No'),
 Text(1674.0, 217.44000000000005, 'gini = 0.278\nsamples = 5\nvalue = [1, 5]\nclass = No'),
 Text(2046.0, 217.44000000000005, 'gini = 0.198\nsamples = 6\nvalue = [1, 8]\nclass = No'),
 Text(2604.0, 652.3200000000002, 'PassengerId <= 496.5\ngini = 0.091\nsamples = 13\nvalue = [20, 1]\nclass = Yes'),
 Text(2418.0, 217.44000000000005, 'gini = 0.0\nsamples = 8\nvalue = [15, 0]\nclass = Yes'),
 Text(2790.0, 217.44000000000005, 'gini = 0.278\nsamples = 5\nvalue = [5, 1]\nclass = Yes'),
 Text(3906.0, 1522.0800000000002, 'Pclass <= 2.0\ngini = 0.344\nsamples = 48\nvalue = [17, 60]\nclass = No'),
 Text(3720.0, 1087.2, 'Survived <= 0.5\ngini = 0.355\nsamples = 41\nvalue = [15, 50]\nclass = No'),
 Text(3348.0, 652.3200000000002, 'PassengerId <= 840.5\ngini = 0.208\nsamples = 22\nvalue = [4, 30]\nclass = No'),
 Text(3162.0, 217.44000000000005, 'gini = 0.133\nsamples = 17\nvalue = [2, 26]\nclass = No'),
 Text(3534.0, 217.44000000000005, 'gini = 0.444\nsamples = 5\nvalue = [2, 4]\nclass = No'),
 Text(4092.0, 652.3200000000002, 'PassengerId <= 794.0\ngini = 0.458\nsamples = 19\nvalue = [11, 20]\nclass = No'),
 Text(3906.0, 217.44000000000005, 'gini = 0.278\nsamples = 10\nvalue = [3, 15]\nclass = No'),
 Text(4278.0, 217.44000000000005, 'gini = 0.473\nsamples = 9\nvalue = [8, 5]\nclass = Yes'),
 Text(4092.0, 1087.2, 'gini = 0.278\nsamples = 7\nvalue = [2, 10]\nclass = No')]
```

