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\train gender.csv")
df
```

Out[2]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	E
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	
...	...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	
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	NaN	1	2	W./C. 6607	23.4500	NaN	
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	
890	891	0	3	Dooley, Mr.	male	32.0	0	0	370376	7.7500	NaN	

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	E
			Patrick								

891 rows × 12 columns

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 [5]:
```
df['Sex'].value_counts()
```

Out[5]:
```
male      577
female    314
Name: Sex, dtype: int64
```

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

Out[6]:
```
male      577
female    314
Name: Sex, dtype: int64
```

In [7]:
```
x=df[['PassengerId', 'Survived', 'Pclass','SibSp',
       'Parch',]]
y=df['Sex']
```

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

```
     PassengerId  Survived  Pclass  \
0              1         0       3
1              2         1       1
2              3         1       3
3              4         1       1
4              5         0       3
..           ...       ...     ...
886          887         0       2
887          888         1       1
888          889         0       3
889          890         1       1
890          891         0       3

                                                 Name     Sex   Age  SibSp  \
0                             Braund, Mr. Owen Harris    male  22.0      1
1    Cumings, Mrs. John Bradley (Florence Briggs Th...  female  38.0      1
2                              Heikkinen, Miss. Laina  female  26.0      0
3        Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0      1
4                            Allen, Mr. William Henry    male  35.0      0
..                                                ...     ...   ...    ...
886                            Montvila, Rev. Juozas    male  27.0      0
887                     Graham, Miss. Margaret Edith  female  19.0      0
888         Johnston, Miss. Catherine Helen "Carrie"  female   NaN      1
889                            Behr, Mr. Karl Howell    male  26.0      0
890                              Dooley, Mr. Patrick    male  32.0      0

     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  13.0000   NaN        S
887      0            112053  30.0000   B42        S
888      2        W./C. 6607  23.4500   NaN        S
889      0            111369  30.0000  C148        C
890      0            370376   7.7500   NaN        Q

[891 rows x 12 columns]
```

In [9]:
```
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]:
```
from sklearn.ensemble import RandomForestClassifier
rfc=RandomForestClassifier()
rfc.fit(x_train,y_train)
```

Out[10]:
```
RandomForestClassifier()
```

In [11]:
```
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]:
```
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]:
```
grid_search.best_score_
```

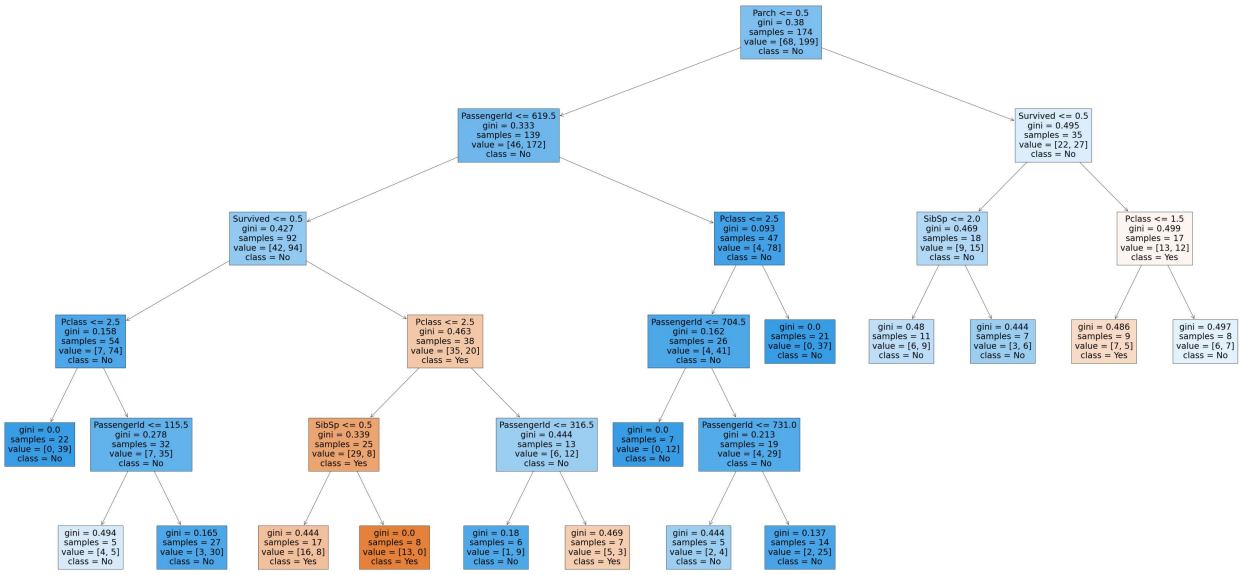
Out[13]:
```
0.8165189092133318
```

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

In [15]:
```
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(2790.0, 1993.2, 'Parch <= 0.5\ngini = 0.38\nsamples = 174\nvalue = [68, 199]\nclas
s = No'),
 Text(1830.24, 1630.8000000000002, 'PassengerId <= 619.5\ngini = 0.333\nsamples = 139\nv
alue = [46, 172]\nclass = No'),
 Text(982.08, 1268.4, 'Survived <= 0.5\ngini = 0.427\nsamples = 92\nvalue = [42, 94]\ncl
ass = No'),
 Text(357.12, 906.0, 'Pclass <= 2.5\ngini = 0.158\nsamples = 54\nvalue = [7, 74]\nclass
= No'),
 Text(178.56, 543.5999999999999, 'gini = 0.0\nsamples = 22\nvalue = [0, 39]\nclass = N
o'),
 Text(535.6800000000001, 543.5999999999999, 'PassengerId <= 115.5\ngini = 0.278\nsamples
= 32\nvalue = [7, 35]\nclass = No'),
 Text(357.12, 181.19999999999982, 'gini = 0.494\nsamples = 5\nvalue = [4, 5]\nclass = N
o'),
 Text(714.24, 181.19999999999982, 'gini = 0.165\nsamples = 27\nvalue = [3, 30]\nclass =
No'),
 Text(1607.04, 906.0, 'Pclass <= 2.5\ngini = 0.463\nsamples = 38\nvalue = [35, 20]\nclas
s = Yes'),
 Text(1249.92, 543.5999999999999, 'SibSp <= 0.5\ngini = 0.339\nsamples = 25\nvalue = [2
9, 8]\nclass = Yes'),
 Text(1071.3600000000001, 181.19999999999982, 'gini = 0.444\nsamples = 17\nvalue = [16,
8]\nclass = Yes'),
 Text(1428.48, 181.19999999999982, 'gini = 0.0\nsamples = 8\nvalue = [13, 0]\nclass = Ye
s'),
 Text(1964.16, 543.5999999999999, 'PassengerId <= 316.5\ngini = 0.444\nsamples = 13\nval
ue = [6, 12]\nclass = No'),
 Text(1785.6, 181.19999999999982, 'gini = 0.18\nsamples = 6\nvalue = [1, 9]\nclass = N
o'),
 Text(2142.7200000000003, 181.19999999999982, 'gini = 0.469\nsamples = 7\nvalue = [5, 3]
\nclass = Yes'),
 Text(2678.4, 1268.4, 'Pclass <= 2.5\ngini = 0.093\nsamples = 47\nvalue = [4, 78]\nclass
= No'),
 Text(2499.84, 906.0, 'PassengerId <= 704.5\ngini = 0.162\nsamples = 26\nvalue = [4, 41]
\nclass = No'),
 Text(2321.28, 543.5999999999999, 'gini = 0.0\nsamples = 7\nvalue = [0, 12]\nclass = N
o'),
 Text(2678.4, 543.5999999999999, 'PassengerId <= 731.0\ngini = 0.213\nsamples = 19\nvalu
e = [4, 29]\nclass = No'),
 Text(2499.84, 181.19999999999982, 'gini = 0.444\nsamples = 5\nvalue = [2, 4]\nclass = N
o'),
 Text(2856.96, 181.19999999999982, 'gini = 0.137\nsamples = 14\nvalue = [2, 25]\nclass =
No'),
```

```
 Text(2856.96, 906.0, 'gini = 0.0\nsamples = 21\nvalue = [0, 37]\nclass = No'),
 Text(3749.76, 1630.8000000000002, 'Survived <= 0.5\ngini = 0.495\nsamples = 35\nvalue =
[22, 27]\nclass = No'),
 Text(3392.64, 1268.4, 'SibSp <= 2.0\ngini = 0.469\nsamples = 18\nvalue = [9, 15]\nclass
= No'),
 Text(3214.08, 906.0, 'gini = 0.48\nsamples = 11\nvalue = [6, 9]\nclass = No'),
 Text(3571.2, 906.0, 'gini = 0.444\nsamples = 7\nvalue = [3, 6]\nclass = No'),
 Text(4106.88, 1268.4, 'Pclass <= 1.5\ngini = 0.499\nsamples = 17\nvalue = [13, 12]\ncla
ss = Yes'),
 Text(3928.32, 906.0, 'gini = 0.486\nsamples = 9\nvalue = [7, 5]\nclass = Yes'),
 Text(4285.4400000000005, 906.0, 'gini = 0.497\nsamples = 8\nvalue = [6, 7]\nclass = N
o')]
```



In [ ]: