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(r"C2_train.gender_submission.csv")
df
```

Out[2]:

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

891 rows × 12 columns

In [3]:

```
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 [4]:

```
df=df.drop('Cabin',axis=1)
```

In [5]:

```
df=df.dropna()
```

In [6]:

```
df.isnull().sum()
```

Out[6]:

```
PassengerId    0
Survived       0
Pclass         0
Name           0
Sex            0
Age            0
SibSp          0
Parch          0
Ticket         0
Fare           0
Embarked       0
dtype: int64
```

In [7]:
```
df.describe()
```

Out[7]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	712.000000	712.000000	712.000000	712.000000	712.000000	712.000000	712.000000
mean	448.589888	0.404494	2.240169	29.642093	0.514045	0.432584	34.567251
std	258.683191	0.491139	0.836854	14.492933	0.930692	0.854181	52.938648
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	222.750000	0.000000	1.000000	20.000000	0.000000	0.000000	8.050000
50%	445.000000	0.000000	2.000000	28.000000	0.000000	0.000000	15.645850
75%	677.250000	1.000000	3.000000	38.000000	1.000000	1.000000	33.000000
max	891.000000	1.000000	3.000000	80.000000	5.000000	6.000000	512.329200

In [8]:
```
df["Survived"].value_counts()
```

Out[8]:
```
0    424
1    288
Name: Survived, dtype: int64
```

In [9]:
```
df1=df[['PassengerId','Survived','Pclass','Age','SibSp','Parch','Fare']]
```

In [10]:
```
x=df1.drop("Survived",axis=1)
y=df1["Survived"]
```

In [11]:
```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,train_size=0.70)
```

In [12]:
```
from sklearn.ensemble import RandomForestClassifier
rfc=RandomForestClassifier()
rfc.fit(x_train,y_train)
```

Out[12]:
```
RandomForestClassifier()
```

In [13]:
```
parameters={'max_depth':[1,2,3,4,5],
            'min_samples_leaf':[5,10,15,20,25],
            'n_estimators':[10,20,30,40,50]}
```

In [14]:
```
from sklearn.model_selection import GridSearchCV
grid_search=GridSearchCV(estimator=rfc,param_grid=parameters,cv=2,scoring="accu
grid_search.fit(x_train,y_train)
```

Out[14]:
```
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 [15]:

```
grid_search.best_score_
```

Out[15]:

```
0.7108433734939759
```

In [16]:

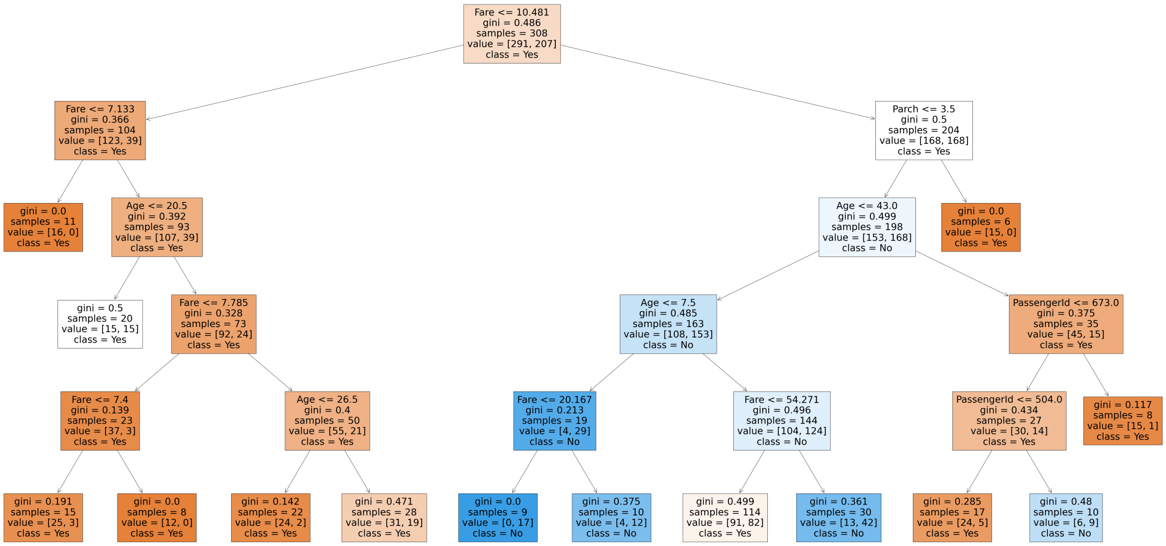
```
parameters=df
```

In [17]:

```
rfc_best=grid_search.best_estimator_
```

In [18]:

```
from sklearn.tree import plot_tree
plt.figure(figsize=(80,40))
plot_tree(rfc_best.estimators_[5],feature_names=x.columns,class_names=['Yes','N
Out[18]: [Text(1966.2857142857144, 1993.2, 'Fare <= 10.481\ngini = 0.486\nsamples = 30
         8\nvalue = [291, 207]\nclass = Yes'),
          Text(425.14285714285717, 1630.8000000000002, 'Fare <= 7.133\ngini = 0.366\ns
         amples = 104\nvalue = [123, 39]\nclass = Yes'),
          Text(212.57142857142858, 1268.4, 'gini = 0.0\nsamples = 11\nvalue = [16, 0]
         \nclass = Yes'),
          Text(637.7142857142858, 1268.4, 'Age <= 20.5\ngini = 0.392\nsamples = 93\nva
         lue = [107, 39]\nclass = Yes'),
          Text(425.14285714285717, 906.0, 'gini = 0.5\nsamples = 20\nvalue = [15, 15]
         \nclass = Yes'),
          Text(850.2857142857143, 906.0, 'Fare <= 7.785\ngini = 0.328\nsamples = 73\nv
         alue = [92, 24]\nclass = Yes'),
          Text(425.14285714285717, 543.5999999999999, 'Fare <= 7.4\ngini = 0.139\nsamp
         les = 23\nvalue = [37, 3]\nclass = Yes'),
          Text(212.57142857142858, 181.19999999999982, 'gini = 0.191\nsamples = 15\nva
         lue = [25, 3]\nclass = Yes'),
          Text(637.7142857142858, 181.19999999999982, 'gini = 0.0\nsamples = 8\nvalue
         = [12, 0]\nclass = Yes'),
          Text(1275.4285714285716, 543.5999999999999, 'Age <= 26.5\ngini = 0.4\nsample
         s = 50\nvalue = [55, 21]\nclass = Yes'),
          Text(1062.857142857143, 181.19999999999982, 'gini = 0.142\nsamples = 22\nval
         ue = [24, 2]\nclass = Yes'),
          Text(1488.0, 181.19999999999982, 'gini = 0.471\nsamples = 28\nvalue = [31, 1
         9]\nclass = Yes'),
          Text(3507.4285714285716, 1630.8000000000002, 'Parch <= 3.5\ngini = 0.5\nsamp
         les = 204\nvalue = [168, 168]\nclass = Yes'),
          Text(3294.857142857143, 1268.4, 'Age <= 43.0\ngini = 0.499\nsamples = 198\nv
         alue = [153, 168]\nclass = No'),
          Text(2550.857142857143, 906.0, 'Age <= 7.5\ngini = 0.485\nsamples = 163\nval
         ue = [108, 153]\nclass = No'),
          Text(2125.714285714286, 543.5999999999999, 'Fare <= 20.167\ngini = 0.213\nsa
         mples = 19\nvalue = [4, 29]\nclass = No'),
          Text(1913.1428571428573, 181.19999999999982, 'gini = 0.0\nsamples = 9\nvalue
         = [0, 17]\nclass = No'),
          Text(2338.285714285714, 181.19999999999982, 'gini = 0.375\nsamples = 10\nval
         ue = [4, 12]\nclass = No'),
          Text(2976.0, 543.5999999999999, 'Fare <= 54.271\ngini = 0.496\nsamples = 144
         \nvalue = [104, 124]\nclass = No'),
          Text(2763.4285714285716, 181.19999999999982, 'gini = 0.499\nsamples = 114\nv
         alue = [91, 82]\nclass = Yes'),
          Text(3188.571428571429, 181.19999999999982, 'gini = 0.361\nsamples = 30\nval
         ue = [13, 42]\nclass = No'),
          Text(4038.857142857143, 906.0, 'PassengerId <= 673.0\ngini = 0.375\nsamples
         = 35\nvalue = [45, 15]\nclass = Yes'),
          Text(3826.2857142857147, 543.5999999999999, 'PassengerId <= 504.0\ngini = 0.
         434\nsamples = 27\nvalue = [30, 14]\nclass = Yes'),
          Text(3613.714285714286, 181.19999999999982, 'gini = 0.285\nsamples = 17\nval
         ue = [24, 5]\nclass = Yes'),
          Text(4038.857142857143, 181.19999999999982, 'gini = 0.48\nsamples = 10\nvalu
         e = [6, 9]\nclass = No'),
          Text(4251.428571428572, 543.5999999999999, 'gini = 0.117\nsamples = 8\nvalue
         = [15, 1]\nclass = Yes'),
          Text(3720.0, 1268.4, 'gini = 0.0\nsamples = 6\nvalue = [15, 0]\nclass = Ye
         s')]
```


Fare <= 10.481
gini = 0.486
samples = 308
value = [291, 207]
class = Yes
Fare <= 7.133
gini = 0.366
samples = 104
value = [123, 39]
class = Yes
Parch <= 3.5
gini = 0.5
samples = 204
value = [168, 168]
class = Yes
gini = 0.0
samples = 11
value = [16, 0]
class = Yes
Age <= 20.5
gini = 0.392
samples = 93
value = [107, 39]
class = Yes
Age <= 43.0
gini = 0.499
samples = 198
value = [153, 168]
class = No
gini = 0.0
samples = 6
value = [15, 0]
class = Yes
gini = 0.5
samples = 20
value = [15, 15]
class = Yes
Fare <= 7.785
gini = 0.328
samples = 73
value = [92, 24]
class = Yes
Age <= 7.5
gini = 0.485
samples = 163
value = [108, 153]
class = No
PassengerId <= 673.0
gini = 0.375
samples = 35
value = [45, 15]
class = Yes
Fare <= 7.4
gini = 0.139
samples = 23
value = [37, 3]
class = Yes
Age <= 26.5
gini = 0.4
samples = 50
value = [55, 21]
class = Yes
Fare <= 20.167
gini = 0.213
samples = 19
value = [4, 29]
class = No
Fare <= 54.271
gini = 0.496
samples = 144
value = [104, 124]
class = No
PassengerId <= 504.0
gini = 0.434
samples = 27
value = [30, 14]
class = Yes
gini = 0.117
samples = 8
value = [15, 1]
class = Yes
gini = 0.191
samples = 15
value = [25, 3]
class = Yes
gini = 0.0
samples = 8
value = [12, 0]
class = Yes
gini = 0.142
samples = 22
value = [24, 2]
class = Yes
gini = 0.471
samples = 28
value = [31, 19]
class = Yes
gini = 0.0
samples = 9
value = [0, 17]
class = No
gini = 0.375
samples = 10
value = [4, 12]
class = No
gini = 0.499
samples = 114
value = [91, 82]
class = Yes
gini = 0.361
samples = 30
value = [13, 42]
class = No
gini = 0.285
samples = 17
value = [24, 5]
class = Yes
gini = 0.48
samples = 10
value = [6, 9]
class = No

In [ ]: