

```
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	Cabin
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="accuracy")
         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')
```

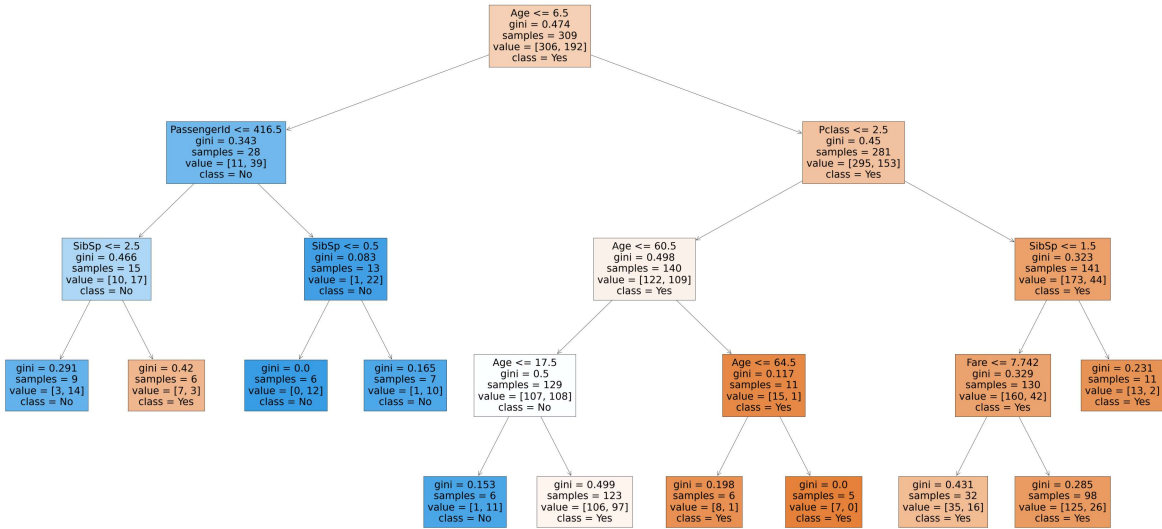
```
In [15]: grid_search.best_score_
```

```
Out[15]: 0.6967871485943775
```

```
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','No'])
```

```
Out[18]: [Text(2064.6, 1956.96, 'Age <= 6.5\ngini = 0.474\nsamples = 309\nvalue = [30
6, 192]\nclasse = Yes'),
Text(892.8, 1522.0800000000002, 'PassengerId <= 416.5\ngini = 0.343\nsamples
= 28\nvalue = [11, 39]\nclasse = No'),
Text(446.4, 1087.2, 'SibSp <= 2.5\ngini = 0.466\nsamples = 15\nvalue = [10,
17]\nclasse = No'),
Text(223.2, 652.32000000000002, 'gini = 0.291\nsamples = 9\nvalue = [3, 14]\n
classe = No'),
Text(669.5999999999999, 652.32000000000002, 'gini = 0.42\nsamples = 6\nvalue
= [7, 3]\nclasse = Yes'),
Text(1339.1999999999998, 1087.2, 'SibSp <= 0.5\ngini = 0.083\nsamples = 13\n
value = [1, 22]\nclasse = No'),
Text(1116.0, 652.32000000000002, 'gini = 0.0\nsamples = 6\nvalue = [0, 12]\nc
lasse = No'),
Text(1562.3999999999999, 652.32000000000002, 'gini = 0.165\nsamples = 7\nvalu
e = [1, 10]\nclasse = No'),
Text(3236.3999999999996, 1522.0800000000002, 'Pclass <= 2.5\ngini = 0.45\nsa
mples = 281\nvalue = [295, 153]\nclasse = Yes'),
Text(2455.2, 1087.2, 'Age <= 60.5\ngini = 0.498\nsamples = 140\nvalue = [12
2, 109]\nclasse = Yes'),
Text(2008.8, 652.32000000000002, 'Age <= 17.5\ngini = 0.5\nsamples = 129\nval
ue = [107, 108]\nclasse = No'),
Text(1785.6, 217.44000000000005, 'gini = 0.153\nsamples = 6\nvalue = [1, 11]
\nclasse = No'),
Text(2232.0, 217.44000000000005, 'gini = 0.499\nsamples = 123\nvalue = [106,
97]\nclasse = Yes'),
Text(2901.6, 652.32000000000002, 'Age <= 64.5\ngini = 0.117\nsamples = 11\nva
lue = [15, 1]\nclasse = Yes'),
Text(2678.3999999999996, 217.44000000000005, 'gini = 0.198\nsamples = 6\nval
ue = [8, 1]\nclasse = Yes'),
Text(3124.7999999999997, 217.44000000000005, 'gini = 0.0\nsamples = 5\nvalue
= [7, 0]\nclasse = Yes'),
Text(4017.6, 1087.2, 'SibSp <= 1.5\ngini = 0.323\nsamples = 141\nvalue = [17
3, 44]\nclasse = Yes'),
Text(3794.3999999999996, 652.32000000000002, 'Fare <= 7.742\ngini = 0.329\nsa
mples = 130\nvalue = [160, 42]\nclasse = Yes'),
Text(3571.2, 217.44000000000005, 'gini = 0.431\nsamples = 32\nvalue = [35, 1
6]\nclasse = Yes'),
Text(4017.6, 217.44000000000005, 'gini = 0.285\nsamples = 98\nvalue = [125,
26]\nclasse = Yes'),
Text(4240.8, 652.32000000000002, 'gini = 0.231\nsamples = 11\nvalue = [13, 2]
\nclasse = Yes')]
```



In []: