```
In [81]: import numpy as np
         import pandas as pd
         import matplotlib.pyplot as plt
```

```
In [97]: df=pd.read_csv(r"C:\USERS\user\Downloads\C2_train.gender_submission - C2_train
```

Out[97]:

	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 [98]:

Out[98]: Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp',
       'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked'],
      dtype='object')

In [99]: ```df=df.head(10)```

Out[99]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Ca
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	N
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	N
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	N
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4583	N
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625	
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.0750	N
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1333	N
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0708	N

In [108]:
```python
a=df[['PassengerId','Survived','Pclass','SibSp','Parch','Fare','Embarked']]
```

Out[108]:

	PassengerId	Survived	Pclass	SibSp	Parch	Fare	Embarked
0	1	0	3	1	0	7.2500	S
1	2	1	1	1	0	71.2833	C
2	3	1	3	0	0	7.9250	S
3	4	1	1	1	0	53.1000	S
4	5	0	3	0	0	8.0500	S
5	6	0	3	0	0	8.4583	Q
6	7	0	1	0	0	51.8625	S
7	8	0	3	3	1	21.0750	S
8	9	1	3	0	2	11.1333	S
9	10	1	2	1	0	30.0708	C

In [109]:
```python
a['Embarked'].value_counts()
```

Out[109]:
```
S    7
C    2
Q    1
Name: Embarked, dtype: int64
```

In [110]:
```python
x=a.drop('Embarked',axis=1)
```

In [111]:
```python
g1={"Embarked":{'S':1,'C':2,'Q':3}}
a=a.replace(g1)
print(a)
```
```
   PassengerId  Survived  Pclass  SibSp  Parch      Fare  Embarked
0            1         0       3      1      0    7.2500         1
1            2         1       1      1      0   71.2833         2
2            3         1       3      0      0    7.9250         1
3            4         1       1      1      0   53.1000         1
4            5         0       3      0      0    8.0500         1
5            6         0       3      0      0    8.4583         3
6            7         0       1      0      0   51.8625         1
7            8         0       3      3      1   21.0750         1
8            9         1       3      0      2   11.1333         1
9           10         1       2      1      0   30.0708         2
```

In [112]:
```python
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 [113]:
```python
from sklearn.ensemble import RandomForestClassifier

rfc=RandomForestClassifier()
rfc.fit(x_train,y_train)
```

Out[113]: RandomForestClassifier()

```
In [114]: parameters={'max_depth':[1,2,3,4,5],
                       'min_samples_leaf':[5,10,15,20,25],
```

```
In [115]: from sklearn.model_selection import GridSearchCV

          grid_search=GridSearchCV(estimator=rfc,param_grid=parameters,cv=2,scoring="acc
```

```
C:\ProgramData\Anaconda3\lib\site-packages\sklearn\model_selection\_split.py:
666: UserWarning: The least populated class in y has only 1 members, which is
less than n_splits=2.
  warnings.warn(("The least populated class in y has only %d"
```

```
Out[115]: 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 [116]:
```

```
Out[116]: 0.875
```

```
In [117]:
```

```
In [118]: from sklearn.tree import plot_tree

          plt.figure(figsize=(80,40))
```

```
Out[118]: [Text(2232.0, 1087.2, 'gini = 0.245\nsamples = 5\nvalue = [1, 6]\nclass = No
          ')]
```



```
In [ ]:
```