

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

```
In [82]: df=pd.read_csv(r"C:\USERS\user\Downloads\C2_test.gender_submission - C2_test.g
```

```
Out[82]:
```

	PassengerId	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
0	892	3	Kelly, Mr. James	male	34.5	0	0	330911	7.8292	NaN
1	893	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.0000	NaN
2	894	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.6875	NaN
3	895	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.6625	NaN
4	896	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298	12.2875	NaN
...
413	1305	3	Spector, Mr. Woolf	male	NaN	0	0	A.5. 3236	8.0500	NaN
414	1306	1	Oliva y Ocana, Dona. Fermina	female	39.0	0	0	PC 17758	108.9000	C105
415	1307	3	Saether, Mr. Simon Sivertsen	male	38.5	0	0	SOTON/O.Q. 3101262	7.2500	NaN
416	1308	3	Ware, Mr. Frederick	male	NaN	0	0	359309	8.0500	NaN
417	1309	3	Peter, Master. Michael J	male	NaN	1	1	2668	22.3583	NaN

418 rows × 11 columns

```
In [83]:
```

```
Out[83]: Index(['PassengerId', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp', 'Parch',
                'Ticket', 'Fare', 'Cabin', 'Embarked'],
                dtype='object')
```

```
In [84]: df=df.head(10)
```

```
Out[84]:
```

	PassengerId	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	892	3	Kelly, Mr. James	male	34.5	0	0	330911	7.8292	NaN	
1	893	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.0000	NaN	
2	894	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.6875	NaN	
3	895	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.6625	NaN	
4	896	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298	12.2875	NaN	
5	897	3	Svensson, Mr. Johan Cervin	male	14.0	0	0	7538	9.2250	NaN	
6	898	3	Connolly, Miss. Kate	female	30.0	0	0	330972	7.6292	NaN	
7	899	2	Caldwell, Mr. Albert Francis	male	26.0	1	1	248738	29.0000	NaN	
8	900	3	Abraham, Mrs. Joseph (Sophie Halaut Easu)	female	18.0	0	0	2657	7.2292	NaN	
9	901	3	Davies, Mr. John Samuel	male	21.0	2	0	A/4 48871	24.1500	NaN	

```
In [85]: a=df[['PassengerId','Pclass','Age','SibSp','Parch','Fare','Embarked']]
```

```
Out[85]:
```

	PassengerId	Pclass	Age	SibSp	Parch	Fare	Embarked
0	892	3	34.5	0	0	7.8292	Q
1	893	3	47.0	1	0	7.0000	S
2	894	2	62.0	0	0	9.6875	Q
3	895	3	27.0	0	0	8.6625	S
4	896	3	22.0	1	1	12.2875	S
5	897	3	14.0	0	0	9.2250	S
6	898	3	30.0	0	0	7.6292	Q
7	899	2	26.0	1	1	29.0000	S
8	900	3	18.0	0	0	7.2292	C
9	901	3	21.0	2	0	24.1500	S

```
In [86]:
```

```
Out[86]: S    6
         Q    3
         C    1
         Name: Embarked, dtype: int64
```

```
In [87]: x=a.drop('Embarked',axis=1)
```

```
In [89]: g1={"Embarked":{"S":1,'C':2,'Q':3}}
         a=a.replace(g1)
```

	PassengerId	Pclass	Age	SibSp	Parch	Fare	Embarked
0	892	3	34.5	0	0	7.8292	3
1	893	3	47.0	1	0	7.0000	1
2	894	2	62.0	0	0	9.6875	3
3	895	3	27.0	0	0	8.6625	1
4	896	3	22.0	1	1	12.2875	1
5	897	3	14.0	0	0	9.2250	1
6	898	3	30.0	0	0	7.6292	3
7	899	2	26.0	1	1	29.0000	1
8	900	3	18.0	0	0	7.2292	2
9	901	3	21.0	2	0	24.1500	1

```
In [90]: from sklearn.model_selection import train_test_split
```

```
In [91]: from sklearn.ensemble import RandomForestClassifier
```

```
rfc=RandomForestClassifier()
```

```
Out[91]: RandomForestClassifier()
```

```
In [92]: parameters={'max_depth':[1,2,3,4,5],
                    'min_samples_leaf':[5,10,15,20,25],
                    'n_estimators':[10,20,30,40,50]}
```

```
In [93]: 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[93]: 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 [94]:
```

```
Out[94]: 0.5833333333333333
```

```
In [95]:
```

```
In [96]: from sklearn.tree import plot_tree

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

```
Out[96]: [Text(2232.0, 1087.2, 'gini = 0.49\nsamples = 3\nvalue = [4, 0, 3]\nnclass = Y
es')]
```

gini = 0.49
samples = 3
value = [4, 0, 3]
class = Yes

```
In [ ]:
```

