```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:
```python
df=pd.read_csv(r"C:\Users\user\Downloads\C2_test.gender_submission - C2_test.gend
df
```

Out[2]:

	PassengerId	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	E
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 [3]:
```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 418 entries, 0 to 417
Data columns (total 11 columns):
 #   Column       Non-Null Count   Dtype
---  ------       --------------   -----
 0   PassengerId  418 non-null     int64
 1   Pclass       418 non-null     int64
 2   Name         418 non-null     object
 3   Sex          418 non-null     object
 4   Age          332 non-null     float64
 5   SibSp        418 non-null     int64
 6   Parch        418 non-null     int64
 7   Ticket       418 non-null     object
 8   Fare         417 non-null     float64
 9   Cabin        91 non-null      object
 10  Embarked     418 non-null     object
dtypes: float64(2), int64(4), object(5)
memory usage: 36.0+ KB
```

In [4]:
```python
df['Embarked'].value_counts()
```

Out[4]:
```
S    270
C    102
Q     46
Name: Embarked, dtype: int64
```

In [5]:
```python
df1=df[['Embarked','PassengerId','Pclass','SibSp','Parch']]
```

In [6]:
```python
x=df1.drop('Embarked',axis=1)
y=df1['Embarked']
```

```python
In [7]: g1={"S":{'S':1,"C":2,"Q":3}}
        df1=df1.replace(g1)
        print(df)
```

```
     PassengerId  Pclass                                        Name  \
0            892       3                            Kelly, Mr. James
1            893       3            Wilkes, Mrs. James (Ellen Needs)
2            894       2                   Myles, Mr. Thomas Francis
3            895       3                            Wirz, Mr. Albert
4            896       3  Hirvonen, Mrs. Alexander (Helga E Lindqvist)
..           ...     ...                                         ...
413         1305       3                          Spector, Mr. Woolf
414         1306       1                Oliva y Ocana, Dona. Fermina
415         1307       3                Saether, Mr. Simon Sivertsen
416         1308       3                       Ware, Mr. Frederick
417         1309       3                     Peter, Master. Michael J

        Sex   Age  SibSp  Parch              Ticket      Fare Cabin Embarked
0      male  34.5      0      0              330911    7.8292   NaN        Q
1    female  47.0      1      0              363272    7.0000   NaN        S
2      male  62.0      0      0              240276    9.6875   NaN        Q
3      male  27.0      0      0              315154    8.6625   NaN        S
4    female  22.0      1      1             3101298   12.2875   NaN        S
..      ...   ...    ...    ...                 ...       ...   ...      ...
413    male   NaN      0      0           A.5. 3236    8.0500   NaN        S
414  female  39.0      0      0            PC 17758  108.9000  C105        C
415    male  38.5      0      0  SOTON/O.Q. 3101262    7.2500   NaN        S
416    male   NaN      0      0              359309    8.0500   NaN        S
417    male   NaN      1      1                2668   22.3583   NaN        C

[418 rows x 11 columns]
```

```python
In [8]: from sklearn.model_selection import train_test_split
        x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=45)
```

Random Forest

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

```
Out[9]: RandomForestClassifier()
```

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

In [11]:
```python
from sklearn.model_selection import GridSearchCV

grid_search =  GridSearchCV(estimator=rfc,param_grid=parameters,cv=2,scoring='acc
grid_search.fit(x_train,y_train)
```

Out[11]:
```
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 [12]:
```python
grid_search.best_score_
```

Out[12]:
```
0.6809556667241676
```

In [13]:
```python
rfc_best = grid_search.best_estimator_
```

In [14]:
```python
# drawing decision tree
from sklearn.tree import plot_tree

plt.figure(figsize=(80,40))
plot_tree(rfc_best.estimators_[5],feature_names=x.columns,filled=True)
```
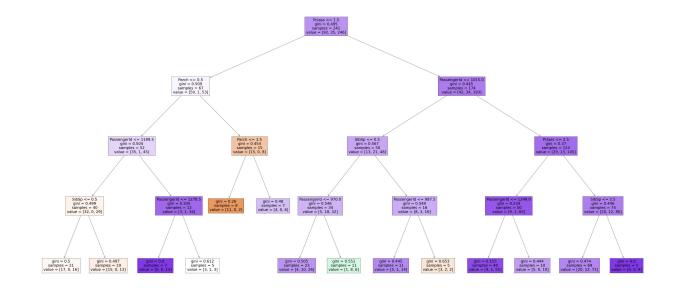
Out[14]: [Text(2103.230769230769, 1956.96, 'Pclass <= 1.5\ngini = 0.495\nsamples = 241\n value = [92, 35, 246]'),
 Text(1116.0, 1522.0800000000002, 'Parch <= 0.5\ngini = 0.509\nsamples = 67\nva lue = [50, 1, 53]'),
 Text(686.7692307692307, 1087.2, 'PassengerId <= 1189.5\ngini = 0.504\nsamples = 52\nvalue = [35, 1, 45]'),
 Text(343.38461538461536, 652.3200000000002, 'SibSp <= 0.5\ngini = 0.499\nsampl es = 40\nvalue = [32, 0, 29]'),
 Text(171.69230769230768, 217.44000000000005, 'gini = 0.5\nsamples = 21\nvalue = [17, 0, 16]'),
 Text(515.0769230769231, 217.44000000000005, 'gini = 0.497\nsamples = 19\nvalue = [15, 0, 13]'),
 Text(1030.1538461538462, 652.3200000000002, 'PassengerId <= 1278.5\ngini = 0.3 35\nsamples = 12\nvalue = [3, 1, 16]'),
 Text(858.4615384615383, 217.44000000000005, 'gini = 0.0\nsamples = 7\nvalue = [0, 0, 13]'),
 Text(1201.8461538461538, 217.44000000000005, 'gini = 0.612\nsamples = 5\nvalue = [3, 1, 3]'),
 Text(1545.230769230769, 1087.2, 'Parch <= 1.5\ngini = 0.454\nsamples = 15\nval ue = [15, 0, 8]'),
 Text(1373.5384615384614, 652.3200000000002, 'gini = 0.26\nsamples = 8\nvalue = [11, 0, 2]'),
 Text(1716.9230769230767, 652.3200000000002, 'gini = 0.48\nsamples = 7\nvalue = [4, 0, 6]'),
 Text(3090.461538461538, 1522.0800000000002, 'PassengerId <= 1015.0\ngini = 0.4 45\nsamples = 174\nvalue = [42, 34, 193]'),
 Text(2403.6923076923076, 1087.2, 'SibSp <= 0.5\ngini = 0.567\nsamples = 50\nva lue = [13, 21, 48]'),
 Text(2060.3076923076924, 652.3200000000002, 'PassengerId <= 970.0\ngini = 0.54 6\nsamples = 34\nvalue = [5, 18, 32]'),
 Text(1888.6153846153845, 217.44000000000005, 'gini = 0.505\nsamples = 23\nvalu e = [4, 10, 26]'),
 Text(2232.0, 217.44000000000005, 'gini = 0.551\nsamples = 11\nvalue = [1, 8, 6]'),
 Text(2747.076923076923, 652.3200000000002, 'PassengerId <= 987.5\ngini = 0.549 \nsamples = 16\nvalue = [8, 3, 16]'),
 Text(2575.3846153846152, 217.44000000000005, 'gini = 0.445\nsamples = 11\nvalu e = [5, 1, 14]'),
 Text(2918.7692307692305, 217.44000000000005, 'gini = 0.653\nsamples = 5\nvalue = [3, 2, 2]'),
 Text(3777.230769230769, 1087.2, 'Pclass <= 2.5\ngini = 0.37\nsamples = 124\nva lue = [29, 13, 145]'),
 Text(3433.8461538461534, 652.3200000000002, 'PassengerId <= 1249.0\ngini = 0.2 34\nsamples = 50\nvalue = [9, 1, 65]'),
 Text(3262.1538461538457, 217.44000000000005, 'gini = 0.155\nsamples = 40\nvalu e = [4, 1, 55]'),
 Text(3605.5384615384614, 217.44000000000005, 'gini = 0.444\nsamples = 10\nvalu e = [5, 0, 10]'),
 Text(4120.615384615385, 652.3200000000002, 'SibSp <= 2.5\ngini = 0.446\nsample s = 74\nvalue = [20, 12, 80]'),

```
  Text(3948.9230769230767, 217.44000000000005, 'gini = 0.474\nsamples = 69\nvalu
e = [20, 12, 71]'),
  Text(4292.307692307692, 217.44000000000005, 'gini = 0.0\nsamples = 5\nvalue =
[0, 0, 9]')]
```


In [ ]: