

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```
In [2]: df=pd.read_csv("C:/Users/Akshay/Desktop/gaurav datascience application/datasets/train.csv")
```

```
In [5]: df
```

```
Out[5]:
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN	Q

891 rows × 12 columns

```
In [ ]:
```

```
In [35]: inputs=df.drop('Survived',axis='columns')
         target=df['Survived']
         target
```

```
Out[35]: 0      0
         1      1
         2      1
         3      1
         4      0
         ..
        886     0
        887     1
        888     0
        889     1
        890     0
        Name: Survived, Length: 891, dtype: int64
```

```
In [36]: from sklearn.preprocessing import LabelEncoder
         le_Pclass=LabelEncoder()
         le_Sex=LabelEncoder()
         le_Age=LabelEncoder()
         le_Fare=LabelEncoder()
```

```
In [37]: inputs['Pclass_n']=le_Pclass.fit_transform(inputs['Pclass'])
inputs['Sex_n']=le_Pclass.fit_transform(inputs['Sex'])
inputs['Age_n']=le_Pclass.fit_transform(inputs['Age'])
inputs['Fare_n']=le_Pclass.fit_transform(inputs['Fare'])
inputs.head()
```

Out[37]:

	PassengerId	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	Pclass_n	Sex_n	Age_n	Fare_n
0	1	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S	2	1	28	18
1	2	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C	0	0	51	207
2	3	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S	2	0	34	41
3	4	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S	0	0	47	189
4	5	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S	2	1	47	43

```
In [38]: inputs_n=inputs.drop(['Pclass', 'Sex', 'Age', 'Fare', 'PassengerId', 'Name', 'SibSp', 'Parch', 'Ticket', 'Cabin', 'Embarked'],axis=1)
inputs_n
```

Out[38]:

	Pclass_n	Sex_n	Age_n	Fare_n
0	2	1	28	18
1	0	0	51	207
2	2	0	34	41
3	0	0	47	189
4	2	1	47	43
...
886	1	1	35	85
887	0	0	24	153
888	2	0	90	131
889	0	1	34	153
890	2	1	42	30

891 rows × 4 columns

```
In [39]: from sklearn import tree
model=tree.DecisionTreeClassifier()
model.fit(inputs_n,target)
```

Out[39]: DecisionTreeClassifier(class_weight=None, criterion='gini', max_depth=None, max_features=None, max_leaf_nodes=None, min_impurity_decrease=0.0, min_impurity_split=None, min_samples_leaf=1, min_samples_split=2, min_weight_fraction_leaf=0.0, presort=False, random_state=None, splitter='best')

```
In [40]: model.score(inputs_n,target)
```

```
Out[40]: 0.9876543209876543
```

```
In [41]: model.predict([[2,1,42,30]])
```

```
Out[41]: array([0], dtype=int64)
```

```
In [ ]:
```

```
In [*]: tree.plot_tree(model)
```

```
Text(317.73333333333335, 77.65714285714284, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),  
Text(322.0, 77.65714285714284, 'X[2] <= 169.0\nngini = 0.26\nsamples = 26\nvalue = [22, 4]'),  
Text(317.73333333333335, 67.30285714285714, 'X[2] <= 94.5\nngini = 0.124\nsamples = 15\nvalue = [14, 1]'),  
Text(315.6, 56.94857142857143, 'X[3] <= 78.5\nngini = 0.444\nsamples = 3\nvalue = [2, 1]'),  
Text(313.46666666666664, 46.59428571428572, 'gini = 0.0\nsamples = 2\nvalue = [2, 0]'),  
Text(317.73333333333335, 46.59428571428572, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),  
Text(319.86666666666667, 56.94857142857143, 'gini = 0.0\nsamples = 12\nvalue = [12, 0]'),  
Text(326.26666666666665, 67.30285714285714, 'X[2] <= 227.5\nngini = 0.397\nsamples = 11\nvalue = [8, 3]'),  
Text(324.13333333333333, 56.94857142857143, 'X[3] <= 101.5\nngini = 0.48\nsamples = 5\nvalue = [2, 3]'),  
Text(322.0, 46.59428571428572, 'gini = 0.0\nsamples = 2\nvalue = [0, 2]'),  
Text(326.26666666666665, 46.59428571428572, 'X[3] <= 119.5\nngini = 0.444\nsamples = 3\nvalue = [2, 1]'),  
Text(324.13333333333333, 36.240000000000001, 'gini = 0.0\nsamples = 2\nvalue = [2, 0]'),  
Text(328.4, 36.240000000000001, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),  
Text(328.4, 56.94857142857143, 'gini = 0.0\nsamples = 6\nvalue = [6, 0]'),  
Text(330.53333333333333, 160.49142857142857, 'X[3] <= 199.0\nngini = 0.43\nsamples = 16\nvalue = [11, 5]'),  
Text(328.4, 150.13714285714286, 'X[2] <= 155.5\nngini = 0.408\nsamples = 7\nvalue = [2, 5]'),  
Text(326.26666666666665, 139.78285714285715, 'X[2] <= 74.5\nngini = 0.48\nsamples = 5\nvalue = [2, 3]'),  
Text(324.13333333333333, 129.42857142857144, 'X[2] <= 39.0\nngini = 0.375\nsamples = 4\nvalue = [1, 3]'),  
Text(322.0, 119.07428571428571, 'X[2] <= 35.0\nngini = 0.5\nsamples = 2\nvalue = [1, 1]'),  
Text(319.86666666666667, 108.72, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
```

```
In [ ]:
```

```
In [ ]:
```

In []:

In []:

In []: