

```
In [1]: #import Libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

```
In [2]: df = pd.read_csv('kyphosis.csv')
```

```
In [3]: df.head()
```

Out[3]:

	Kyphosis	Age	Number	Start
0	absent	71	3	5
1	absent	158	3	14
2	present	128	4	5
3	absent	2	5	1
4	absent	1	4	15

```
In [4]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 81 entries, 0 to 80
Data columns (total 4 columns):
Kyphosis      81 non-null object
Age           81 non-null int64
Number        81 non-null int64
Start         81 non-null int64
dtypes: int64(3), object(1)
memory usage: 2.6+ KB
```

```
In [5]: from sklearn.model_selection import train_test_split
```

```
In [6]: X = df.drop('Kyphosis',axis=1)
y = df['Kyphosis']
```

```
In [7]: X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.30)
```

```
In [8]: from sklearn.tree import DecisionTreeClassifier
```

```
In [9]: dtree = DecisionTreeClassifier()
```

```
In [10]: dtree.fit(X_train,y_train)
```

```
Out[10]: 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 [11]: predictions = dtree.predict(X_test)
```

```
In [12]: predictions
```

```
Out[12]: array(['absent', 'present', 'absent', 'absent', 'present', 'absent',  
               'present', 'absent', 'present', 'present', 'present', 'absent',  
               'absent', 'present', 'absent', 'present', 'absent', 'present',  
               'absent', 'absent', 'present', 'absent', 'present', 'absent',  
               'absent'], dtype=object)
```