

```
In [28]: import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
dr=pd.read_csv(r"C:\Users\ubin1\Downloads\drug200.csv")
dr
```

Out[28]:

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
0	23	F	HIGH	HIGH	25.355	drugY
1	47	M	LOW	HIGH	13.093	drugC
2	47	M	LOW	HIGH	10.114	drugC
3	28	F	NORMAL	HIGH	7.798	drugX
4	61	F	LOW	HIGH	18.043	drugY
...	...	...	...	...	...	...
195	56	F	LOW	HIGH	11.567	drugC
196	16	M	LOW	HIGH	12.006	drugC
197	52	M	NORMAL	HIGH	9.894	drugX
198	23	M	NORMAL	NORMAL	14.020	drugX
199	40	F	LOW	NORMAL	11.349	drugX

200 rows × 6 columns

```
In [29]: dr.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 6 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Age              200 non-null    int64
1   Sex              200 non-null    object
2   BP               200 non-null    object
3   Cholesterol       200 non-null    object
4   Na_to_K          200 non-null    float64
5   Drug             200 non-null    object
dtypes: float64(1), int64(1), object(4)
memory usage: 9.5+ KB
```

```
In [30]: dr['Cholesterol'].value_counts()
```

```
Out[30]: Cholesterol
HIGH      103
NORMAL     97
Name: count, dtype: int64
```

```
In [31]: dr['BP'].value_counts()
```

```
Out[31]: BP
HIGH      77
LOW       64
NORMAL    59
Name: count, dtype: int64
```

```
In [32]: dr['Drug'].value_counts()
```

```
Out[32]: Drug
drugY     91
drugX     54
drugA     23
drugC     16
drugB     16
Name: count, dtype: int64
```

```
In [33]: convert={'BP':{'HIGH':1,'LOW':2,'NORMAL':3}}
dr=dr.replace(convert)
dr
```

```
Out[33]:
```

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
0	23	F	1	HIGH	25.355	drugY
1	47	M	2	HIGH	13.093	drugC
2	47	M	2	HIGH	10.114	drugC
3	28	F	3	HIGH	7.798	drugX
4	61	F	2	HIGH	18.043	drugY
...	...	...	...	...	...	...
195	56	F	2	HIGH	11.567	drugC
196	16	M	2	HIGH	12.006	drugC
197	52	M	3	HIGH	9.894	drugX
198	23	M	3	NORMAL	14.020	drugX
199	40	F	2	NORMAL	11.349	drugX

200 rows × 6 columns

```
In [34]: convert={'Cholesterol':{'HIGH':1, 'NORMAL':2}}
dr=dr.replace(convert)
dr
```

Out[34]:

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
0	23	F	1	1	25.355	drugY
1	47	M	2	1	13.093	drugC
2	47	M	2	1	10.114	drugC
3	28	F	3	1	7.798	drugX
4	61	F	2	1	18.043	drugY
...	...	...	...	...	...	...
195	56	F	2	1	11.567	drugC
196	16	M	2	1	12.006	drugC
197	52	M	3	1	9.894	drugX
198	23	M	3	2	14.020	drugX
199	40	F	2	2	11.349	drugX

```
In [35]: convert={'Sex':{'F':1, 'M':2}}
dr=dr.replace(convert)
dr
```

Out[35]:

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
0	23	1	1	1	25.355	drugY
1	47	2	2	1	13.093	drugC
2	47	2	2	1	10.114	drugC
3	28	1	3	1	7.798	drugX
4	61	1	2	1	18.043	drugY
...	...	...	...	...	...	...
195	56	1	2	1	11.567	drugC
196	16	2	2	1	12.006	drugC
197	52	2	3	1	9.894	drugX
198	23	2	3	2	14.020	drugX
199	40	1	2	2	11.349	drugX

200 rows × 6 columns

```
In [36]: x=['Age','Sex','BP','Cholesterol','Na_to_K']  
y=['Yes','No']  
all_inputs=dr[x]  
all_classes=dr['Drug']
```

```
In [37]: x_train,x_test,y_train,y_test=train_test_split(all_inputs,all_classes,test_size=0.5
```

```
In [38]: dr=DecisionTreeClassifier(random_state=0)
```

```
In [39]: dr.fit(x_train,y_train)
```

```
Out[39]: 

DecisionTreeClassifier



DecisionTreeClassifier(random_state=0)


```

```
In [40]: score=dr.score(x_test,y_test)  
print(score)
```

0.98