In [1]:

```
import pandas as pd
import numpy as np
from sklearn.tree import  DecisionTreeClassifier
from sklearn.model_selection import train_test_split

```

In [2]:

```
df=pd.read_csv(r"C:\Users\MY HOME\Desktop\datascience\drug200.csv")
df
```

Out[2]:

	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 [3]:

```
df.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 [4]:

```
1 df.isna().any()
```

Out[4]:

```
Age            False
Sex            False
BP             False
Cholesterol    False
Na_to_K        False
Drug           False
dtype: bool
```

In [5]:

```
1 df.describe()
```

Out[5]:

	Age	Na_to_K
count	200.000000	200.000000
mean	44.315000	16.084485
std	16.544315	7.223956
min	15.000000	6.269000
25%	31.000000	10.445500
50%	45.000000	13.936500
75%	58.000000	19.380000
max	74.000000	38.247000

In [6]:

```
1 df.shape
```

Out[6]:

```
(200, 6)
```

In [7]:

```
1 df["Cholesterol"].value_counts()
```

Out[7]:

```
Cholesterol
HIGH      103
NORMAL     97
Name: count, dtype: int64
```

In [8]:

```
convert={"Cholesterol":{"HIGH":1,"NORMAL":0}}
df=df.replace(convert)
df
```

Out[8]:

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

200 rows × 6 columns

In [9]:

```
convert={"Sex":{"F":1,"M":0}}
df=df.replace(convert)
df
```

Out[9]:

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

200 rows × 6 columns

In [10]:

```
convert={"BP":{"HIGH":1,"LOW":0,"NORMAL":2}}
df=df.replace(convert)
df
```

Out[10]:

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

200 rows × 6 columns

In [11]:

```
convert={"Drug":{"drugX":0,"drugC":1,"drugY":2}}
df.replace(convert)
df
```

Out[11]:

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

200 rows × 6 columns

In [16]:

```
x=["Age","Sex","BP","Na_to_K","Cholesterol"]
y=["0","1","2"]
all_inputs=df[x]
all_classes=df["Drug"]
x_train,x_test,y_train,y_test=train_test_split(all_inputs,all_classes,train_size=0.5

```

In [13]:

```
clf=DecisionTreeClassifier(random_state=10)
clf.fit(x_train,y_train)
score=clf.score(x_test,y_test)
print(score)
```

```
0.99
```

In [ ]:

```

```