In [1]:
```
import numpy as np
import pandas as pd
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
```

In [2]:
```
df=pd.read_csv(r"C:\Users\HP\Downloads\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]:
```
df['BP'].value_counts()
```

Out[4]:
```
BP
HIGH      77
LOW       64
NORMAL    59
Name: count, dtype: int64
```

In [5]:
```
df['Drug'].value_counts()
```

Out[5]:
```
Drug
drugY    91
drugX    54
drugA    23
drugC    16
drugB    16
Name: count, dtype: int64
```

In [6]:
```
c={"Cholesterol":{"HIGH":2,"NORMAL":1,"LOW":0}}
df=df.replace(c)
df
```

Out[6]:

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

200 rows × 6 columns

In [8]:
```
d={"Drug":{"drugY":2,"drugC":1,"drugX":0}}
df=df.replace(d)
df
```

Out[8]:

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

200 rows × 6 columns

In [11]:
```
x=["Age","Cholesterol","Na_to_K"]
y=['NORMAL','LOW',"HIGH"]
all_inputs=df[x]
all_classes=df["BP"]
```

In [12]:
```
x_train,x_test,y_train,y_test=train_test_split(all_inputs,all_classes,test_siz
clf=DecisionTreeClassifier(random_state=0)
clf.fit(x_train,y_train)
```

Out[12]:

```
DecisionTreeClassifier
DecisionTreeClassifier(random_state=0)
```

In [13]:
```
score=clf.score(x_test,y_test)
print(score)
```

```
0.28
```

In [ ]: