In [65]:
```python
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 [66]:
```python
df=pd.read_csv(r"C:\Users\Lenovo\OneDrive\Desktop\Data Sets\drug200.csv")
df
```

Out[66]:

	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 [67]:
```python
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 [68]: `df['Drug'].value_counts()`

Out[68]:
```
Drug
drugY    91
drugX    54
drugA    23
drugC    16
drugB    16
Name: count, dtype: int64
```

In [69]: `df['Sex'].value_counts()`

Out[69]:
```
Sex
M    104
F     96
Name: count, dtype: int64
```

In [70]: `df['Cholesterol'].value_counts()`

Out[70]:
```
Cholesterol
HIGH      103
NORMAL     97
Name: count, dtype: int64
```

In [71]: `df['BP'].value_counts()`

Out[71]:
```
BP
HIGH      77
LOW       64
NORMAL    59
Name: count, dtype: int64
```

In [72]:
```python
convert={"Cholesterol":{"HIGH":1,"NORMAL":2}}
df=df.replace(convert)
df
```

Out[72]:

	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	2	14.020	drugX
199	40	F	LOW	2	11.349	drugX

200 rows × 6 columns

In [73]:
```python
convert={"BP":{"HIGH":3,"NORMAL":2,"LOW":1}}
df=df.replace(convert)
df
```

Out[73]:

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

200 rows × 6 columns

In [74]:
```python
convert={"Sex":{"M":1,"F":0}}
df=df.replace(convert)
df
```

Out[74]:

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

200 rows × 6 columns

In [75]:
```python
x=["Sex","BP","Cholesterol"]
y=["DrugY","DrugX","DrugA","DrugC","DrugB"]
all_inputs=df[x]
all_classes=df["Drug"]
```

In [76]:
```python
x_train,x_test,y_train,y_test=train_test_split(all_inputs,all_classes,test_size
clf=DecisionTreeClassifier(random_state=0)
clf.fit(x_train,y_train)
```

Out[76]:
```
▼        DecisionTreeClassifier

DecisionTreeClassifier(random_state=0)
```

In [77]:
```python
score=clf.score(x_test,y_test)
print(score)
```

```
0.5
```