

DRUGS USING DECISION TREE

In [3]:

```
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 [4]:

```
df=pd.read_csv(r"C:\Users\RAMADEVI SURIPAKA\Downloads\drug200.csv")
df
```

Out[4]:

	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 [5]:

```
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 [6]:

```
df['BP'].value_counts()
```

Out[6]:

```
BP
HIGH      77
LOW       64
NORMAL    59
Name: count, dtype: int64
```

In [7]:

```
df['Drug'].value_counts()
```

Out[7]:

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

In [8]:

```
converter={"Sex":{"Yes":1, "No":2}}
df=df.replace(converter)
df
```

Out[8]:

	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 [9]:

```
converter={"Sex":{"F":1,"M":2}}
df=df.replace(converter)
df
```

Out[9]:

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

200 rows × 6 columns

In [15]:

```
x=["Na_to_K","Age"]
y=["Yes","No"]
all_inputs=df[x]
all_classes=df["Sex"]
```

In [16]:

```
(x_train,x_test,y_train,y_test)=train_test_split(all_inputs,all_classes,test_size=0.5)
```

In [17]:

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

In [18]:

```
clf.fit(x_train,y_train)
```

Out[18]:

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

In [19]:

```
score=clf.score(x_test,y_test)  
print(score)
```

0.47

In []:

In []: