

```
In [2]: import numpy as np
import pandas as pd
import seaborn as sb
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
```

```
In [3]: df=pd.read_csv(r"C:\Users\pappu\Downloads\drug200.csv")
df
```

```
Out[3]:
```

	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 [10]: df.head()
```

```
Out[10]:
```

	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

In [4]: 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 [9]: df.describe()

Out[9]:

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

Out[6]: BP

HIGH	77
LOW	64
NORMAL	59

Name: count, dtype: int64

In [7]: df['Cholesterol'].value_counts()

Out[7]: Cholesterol

HIGH	103
NORMAL	97

Name: count, dtype: int64

```
In [8]: df['Drug'].value_counts()
```

```
Out[8]: Drug
drugY    91
drugX    54
drugA    23
drugC    16
drugB    16
Name: count, dtype: int64
```

```
In [15]: r={'BP':{'LOW':0,'NORMAL':1,'HIGH':2}}
df=df.replace(r)
df
```

```
Out[15]:
```

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

200 rows × 6 columns

```
In [16]: r={'Cholesterol':{'NORMAL':0,'HIGH':1}}
df=df.replace(r)
df
```

```
Out[16]:
```

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

200 rows × 6 columns

```
In [17]: x=['Age','BP','Cholesterol']
y=['drugY','drugX','drugA','drugC','drugC']
features=df[x]
target=df['Drug']
```

```
In [18]: (x_train,x_test,y_train,y_test)=train_test_split(features,target,test_size=0.2)
```

```
In [19]: dt=DecisionTreeClassifier()
```

```
In [20]: dt.fit(x_train,y_train)
```

```
Out[20]: DecisionTreeClassifier()
```

In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.

On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

```
In [21]: score=dt.score(x_test,y_test)
print(score)
```

0.475

```
In [ ]:
```

