

```
In [1]: import pandas as pd
        from sklearn.tree import DecisionTreeClassifier
```

```
In [2]: df = pd.read_csv(r"C:\Users\hp\Documents\Datasets\drug200.csv")
        df.head()
```

```
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

Checking Missing Values

```
In [3]: df.isnull().sum()
```

```
Out[3]: Age          0
        Sex          0
        BP          0
        Cholesterol  0
        Na_to_K      0
        Drug         0
        dtype: int64
```

```
In [10]: ## Checking Unique Values
         df["Cholesterol"].unique()
```

```
Out[10]: array(['HIGH', 'NORMAL'], dtype=object)
```

```
In [21]: ## Converting Values to numeric
```

```
In [16]: Chloes_dict = {"HIGH":1, "NORMAL":0}
         df["Cholestrol_1"] = df.Cholesterol.map(Chloes_dict)
```

```
In [17]: Sex_Dict = {"M":0, "F":1}
         df["Sex_1"] = df.Sex.map(Sex_Dict)
```

```
In [18]: df["BP"].unique()
```

```
Out[18]: array(['HIGH', 'LOW', 'NORMAL'], dtype=object)
```

```
In [19]: BP_dict = {"HIGH":2, "LOW":0, "NORMAL":1}
         df["BP_1"] = df.BP.map(BP_dict)
```

```
In [20]: df.head()
```

```
Out[20]:
```

	Age	Sex	BP	Cholesterol	Na_to_K	Drug	Cholestrol_1	Sex_1	BP_1
0	23	F	HIGH	HIGH	25.355	DrugY	1	1	2
1	47	M	LOW	HIGH	13.093	drugC	1	0	0
2	47	M	LOW	HIGH	10.114	drugC	1	0	0
3	28	F	NORMAL	HIGH	7.798	drugX	1	1	1
4	61	F	LOW	HIGH	18.043	DrugY	1	1	0

```
In [30]: df1 = df.drop(["BP","Cholesterol","Sex"],axis=1)
```

```
In [31]: ## Splitting Values
X = df1.drop(["Drug"],axis=1)
Y = df1["Drug"]
```

```
In [32]: from sklearn.model_selection import train_test_split
```

```
In [33]: X_train,X_test,Y_train,Y_test = train_test_split(X,Y,test_size=0.2,random_state=1)
```

```
In [34]: Dt = DecisionTreeClassifier()
```

```
In [36]: Dt.fit(X_train,Y_train)
```

```
Out[36]: DecisionTreeClassifier()
```

```
In [39]: Dt.score(X_test,Y_test)
```

```
Out[39]: 1.0
```

```
In [43]: Y_pred = Dt.predict(X_test)
Y_pred
```

```
Out[43]: array(['drugX', 'DrugY', 'drugX', 'drugC', 'DrugY', 'drugX', 'drugX',
        'DrugY', 'DrugY', 'DrugY', 'drugX', 'drugC', 'DrugY', 'DrugY',
        'drugA', 'drugA', 'drugX', 'drugX', 'drugB', 'DrugY', 'drugX',
        'drugX', 'drugX', 'DrugY', 'drugB', 'drugX', 'drugX', 'DrugY',
        'drugC', 'drugX', 'drugC', 'DrugY', 'DrugY', 'DrugY', 'drugA',
        'DrugY', 'drugA', 'DrugY', 'DrugY', 'DrugY'], dtype=object)
```

```
In [41]: from sklearn.metrics import accuracy_score
```

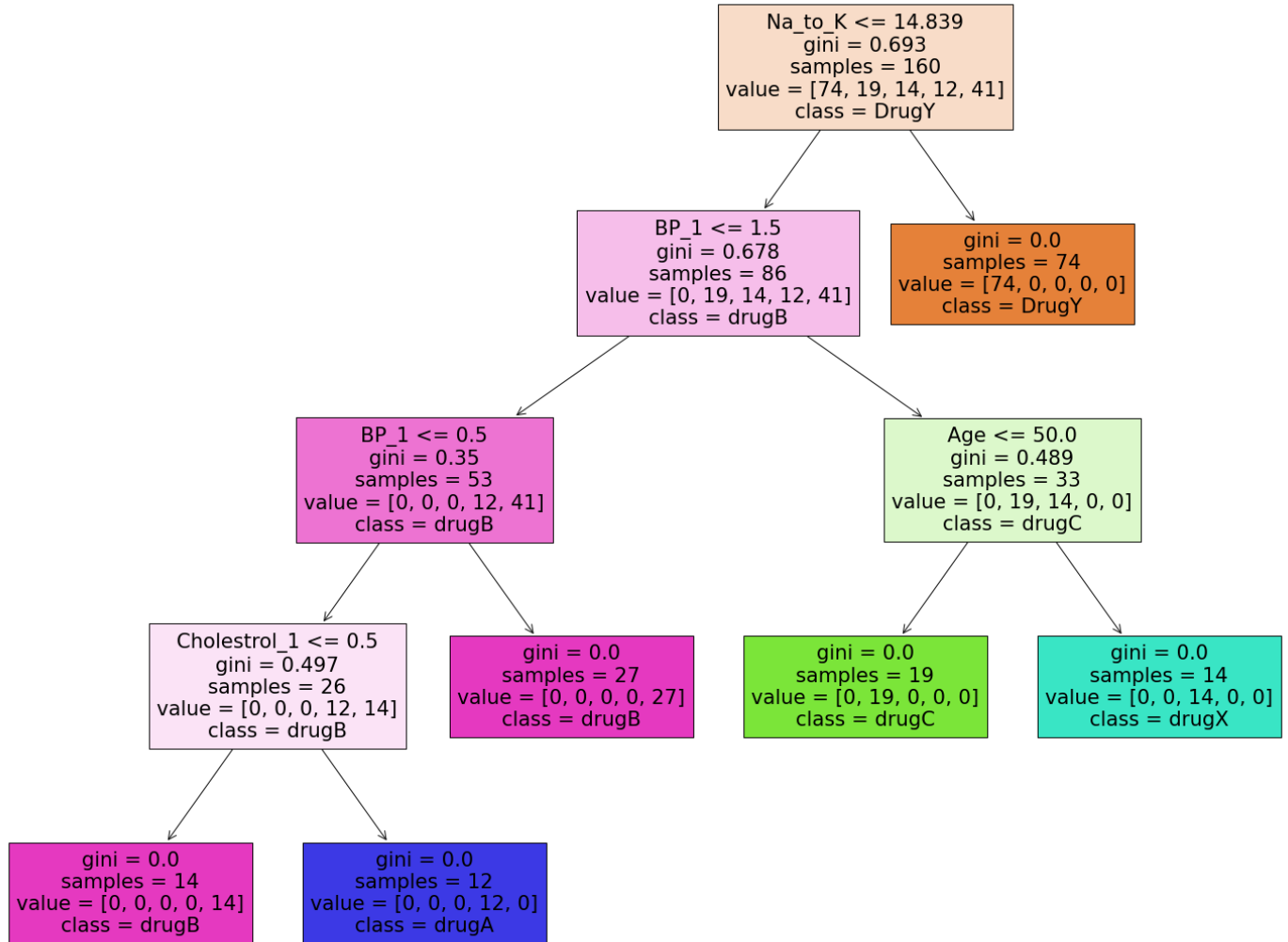
```
In [44]: accuracy_score(Y_pred,Y_test)
```

```
Out[44]: 1.0
```

```
In [58]: clf = DecisionTreeClassifier(random_state = 12)
model = clf.fit(X_train,Y_train)
```

```
from sklearn import tree
import matplotlib.pyplot as plt
```

```
In [61]: fig = plt.figure(figsize=(25,20))
_ = tree.plot_tree(clf,
                  feature_names=X.columns,
                  class_names=Y.unique(),
                  filled=True)
```



```
In [ ]:
```