

```
In [1]: #importing libraries
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.metrics import accuracy_score
from sklearn.tree import plot_tree

import warnings
warnings.filterwarnings("ignore")
```

```
In [2]: #Loading dataset
df = pd.read_csv('iris.csv')
df
```

0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa
...
145	146	6.7	3.0	5.2	2.3	Iris-virginica
146	147	6.3	2.5	5.0	1.9	Iris-virginica
147	148	6.5	3.0	5.2	2.0	Iris-virginica
148	149	6.2	3.4	5.4	2.3	Iris-virginica
149	150	5.9	3.0	5.1	1.8	Iris-virginica

150 rows × 6 columns

150 rows × 6 columns

```
In [3]: df.describe()
```

Out[3]:

	Id	Sepal.LengthCm	Sepal.WidthCm	Petal.LengthCm	Petal.WidthCm
count	150.000000	150.000000	150.000000	150.000000	150.000000
mean	75.500000	5.843333	3.054000	3.758667	1.198667
std	43.445368	0.828066	0.433594	1.764420	0.763161
min	1.000000	4.300000	2.000000	1.000000	0.100000
25%	38.250000	5.100000	2.800000	1.600000	0.300000
50%	75.500000	5.800000	3.000000	4.350000	1.300000
75%	112.750000	6.400000	3.300000	5.100000	1.800000
max	150.000000	7.900000	4.400000	6.900000	2.500000

```
In [4]: #Checking Null Values
df.isnull()
```

Out[4]:

	Id	Sepal.LengthCm	Sepal.WidthCm	Petal.LengthCm	Petal.WidthCm	Species
0	False	False	False	False	False	False
1	False	False	False	False	False	False
2	False	False	False	False	False	False
3	False	False	False	False	False	False
4	False	False	False	False	False	False
...
145	False	False	False	False	False	False
146	False	False	False	False	False	False
147	False	False	False	False	False	False
148	False	False	False	False	False	False
149	False	False	False	False	False	False

150 rows × 6 columns

```
In [5]: df.head()
```

Out[5]:

	Id	Sepal.LengthCm	Sepal.WidthCm	Petal.LengthCm	Petal.WidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa

```
In [7]: #Shape of Dataset rows and columns
df.shape
```

Out[7]: (150, 6)

```
In [9]: #Data Visualization
sns.pairplot(df, hue='Species',palette='coolwarm')
```



```
In [10]: X=df.iloc[:,[0,1,2,3]]
y=df.Species
y
df.shape
```

Out[10]: (150, 6)

```
In [11]: X.shape
```

Out[11]: (150, 4)

```
In [12]: y.shape
```

Out[12]: (150,)

```
In [13]: #Spliting Data

from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y)
```

```
In [14]: X_train.shape
```

Out[14]: (112, 4)

```
In [15]: X_test.shape
```

Out[15]: (38, 4)

```
In [16]: y_train.shape
```

Out[16]: (112,)

```
In [17]: y_test.shape
```

Out[17]: (38,)

```
In [20]: #Building DecisionTree Classifier Model

from sklearn.tree import DecisionTreeClassifier
clf = DecisionTreeClassifier()
clf.fit(X_train, y_train)
```

Out[20]: DecisionTreeClassifier()

```
In [21]: y_pred_train = clf.predict(X_train)
y_pred_test = clf.predict(X_test)
```

```
In [22]: pd.DataFrame(df.Species).value_counts()
```

Out[22]: Species
Iris-setosa 50
Iris-versicolor 50
Iris-virginica 50
dtype: int64

```
In [23]: # Testing the accuracy of model prediction

accuracy_score(y_train, y_pred_train)
```

Out[23]: 1.0

```
In [24]: accuracy_score(y_test, y_pred_test)
```

Out[24]: 0.9473684210526315

```
In [30]: f_n=["Sepal length", "Sepal width", "Petal length", "Petal width"]
c_n=["Setosa","Versicolor","Virginica"]
plot_tree(clf,feature_names=f_n,class_names=c_n, filled=True)
```

Out[30]: [Text(2678.3999999999996, 2265.0, 'Sepal length <= 102.5\ngini = 0.663\nsamples = 112\nvalue = [34, 43, 35]\nnclass = Versicolor'),
Text(1785.6, 1359.0, 'Sepal length <= 50.0\ngini = 0.493\nsamples = 77\nvalue = [34, 43, 0]\nnclass = Versicolor'),
Text(892.8, 453.0, 'gini = 0.0\nsamples = 34\nvalue = [34, 0, 0]\nnclass = Setosa'),
Text(2678.3999999999996, 453.0, 'gini = 0.0\nsamples = 43\nvalue = [0, 43, 0]\nnclass = Versicolor'),
Text(3571.2, 1359.0, 'gini = 0.0\nsamples = 35\nvalue = [0, 0, 35]\nnclass = Virginica')]

Sepal length <= 102.5
gini = 0.663
samples = 112
value = [34, 43, 35]
class = Versicolor

Sepal length <= 50.0
gini = 0.493
samples = 77
value = [34, 43, 0]
class = Versicolor

gini = 0.0
samples = 35
value = [0, 0, 35]
class = Virginica

gini = 0.0
samples = 34
value = [34, 0, 0]
class = Setosa

gini = 0.0
samples = 43
value = [0, 43, 0]
class = Versicolor

```
In [ ]:
```