

LetsGrowMore DataScience Internship

Ashmita Bhattacharjee

mimibhattaworkpurpose0408@gmail.com

```
In [20]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn import tree
```

```
In [2]: df = pd.read_csv('Iris.csv')
df.head()
```

Out[2]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	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 [3]: df.describe()
```

Out[3]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
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]: df.info
```

Out[4]:

<bound method DataFrame.info of		Id	SepalLengthCm	SepalWidthCm	PetalLengthCm
PetalWidthCm \					
0	1	5.1	3.5	1.4	0.2
1	2	4.9	3.0	1.4	0.2
2	3	4.7	3.2	1.3	0.2
3	4	4.6	3.1	1.5	0.2
4	5	5.0	3.6	1.4	0.2
⋮	⋮	⋮	⋮	⋮	⋮
145	146	6.7	3.0	5.2	2.3

146	147	6.3	2.5	5.0	1.9
147	148	6.5	3.0	5.2	2.0
148	149	6.2	3.4	5.4	2.3
149	150	5.9	3.0	5.1	1.8

	Species
0	Iris-setosa
1	Iris-setosa
2	Iris-setosa
3	Iris-setosa
4	Iris-setosa
..	...
145	Iris-virginica
146	Iris-virginica
147	Iris-virginica
148	Iris-virginica
149	Iris-virginica

[150 rows x 6 columns]>

```
In [5]: Iris = df.drop(['Id'], axis = 1)
```

```
In [6]: Iris.head()
```

```
Out[6]:
```

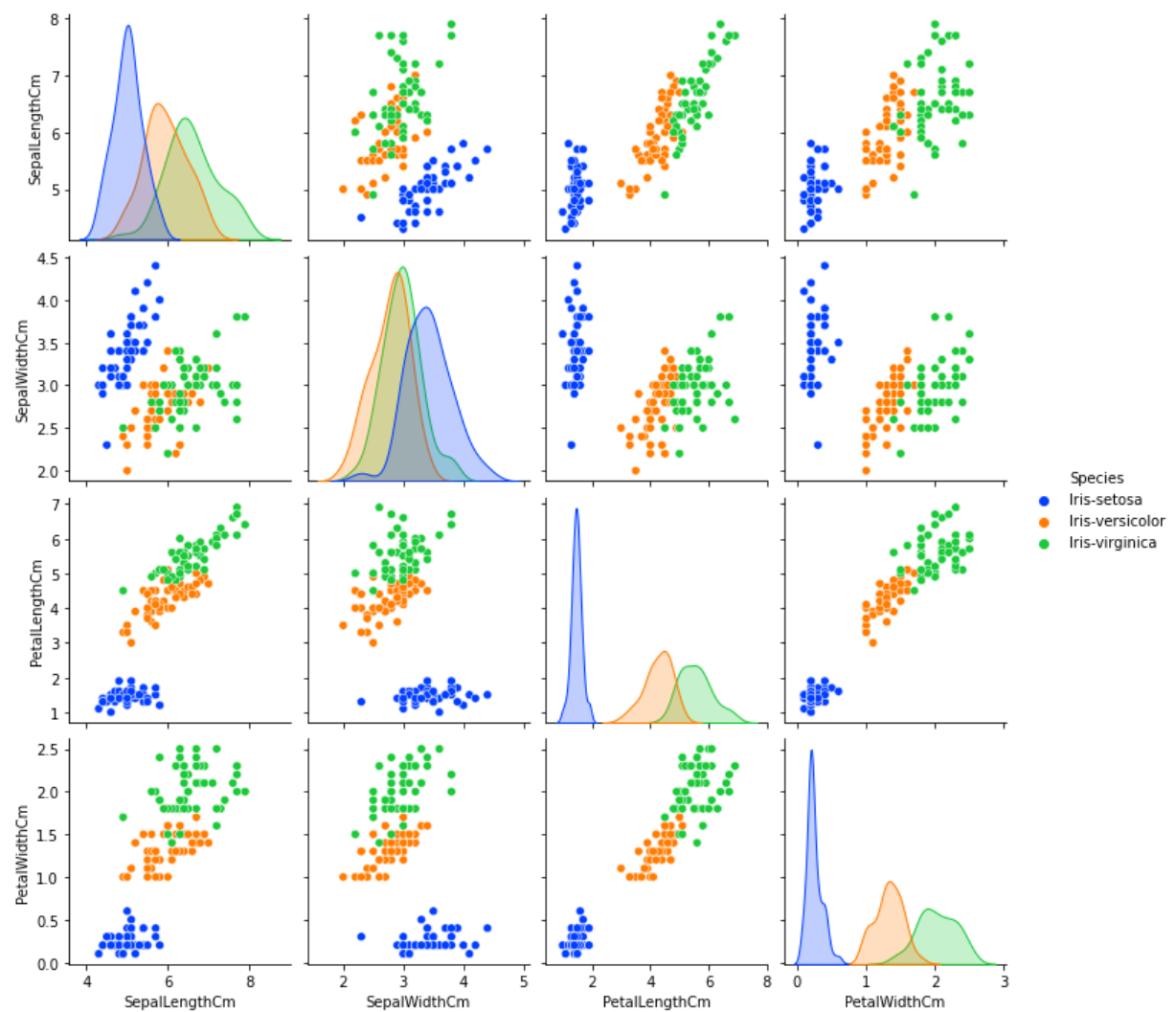
	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa

```
In [7]: Iris.Species.value_counts()
```

```
Out[7]: Iris-virginica    50
Iris-versicolor    50
Iris-setosa    50
Name: Species, dtype: int64
```

```
In [8]: import seaborn as sns
sns.pairplot(Iris, hue = 'Species', palette = 'bright')
```

```
Out[8]: <seaborn.axisgrid.PairGrid at 0x1ef7b99a6a0>
```



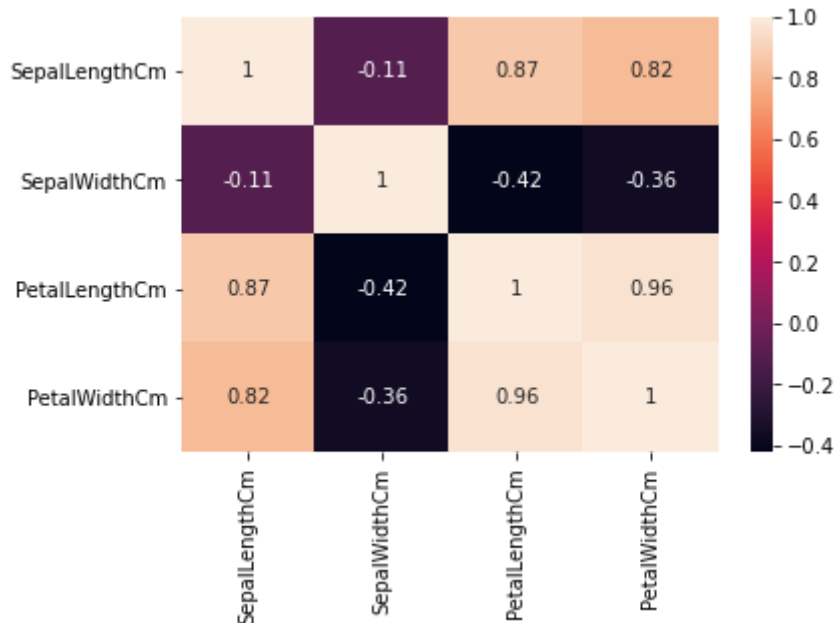
```
In [9]: corr = Iris.corr()
        corr
```

Out[9]:

	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
SepalLengthCm	1.000000	-0.109369	0.871754	0.817954
SepalWidthCm	-0.109369	1.000000	-0.420516	-0.356544
PetalLengthCm	0.871754	-0.420516	1.000000	0.962757
PetalWidthCm	0.817954	-0.356544	0.962757	1.000000

```
In [10]: sns.heatmap(corr, annot=True)
```

```
Out[10]: <AxesSubplot:>
```



```
In [11]: from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
Iris['Species'] = le.fit_transform(Iris['Species'])
Iris.head()
```

```
Out[11]:
```

	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	5.1	3.5	1.4	0.2	0
1	4.9	3.0	1.4	0.2	0
2	4.7	3.2	1.3	0.2	0
3	4.6	3.1	1.5	0.2	0
4	5.0	3.6	1.4	0.2	0

```
In [12]: Iris.Species.value_counts()
```

```
Out[12]: 0    50
1     50
2     50
Name: Species, dtype: int64
```

```
In [13]: from sklearn.model_selection import train_test_split
x = Iris.drop('Species', axis = 1)
y = Iris.Species
x_train,x_test, y_train, y_test = train_test_split(x, y, test_size = 0.30)
```

```
In [14]: from sklearn.tree import DecisionTreeClassifier
from sklearn import metrics
dtree = DecisionTreeClassifier()
dtree.fit(x_train,y_train)
pred = dtree.predict(x_test)
```

```
In [15]: print('The Accuracy of our model is:', metrics.accuracy_score(pred, y_test))
```

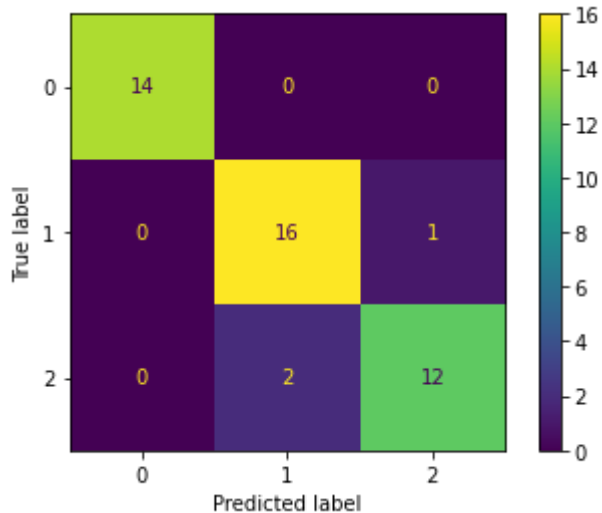
The Accuracy of our model is: 0.9333333333333333

```
In [16]: import sklearn
sklearn.metrics.confusion_matrix(y_test,pred)
```

```
Out[16]: array([[14,  0,  0],
               [ 0, 16,  1],
               [ 0,  2, 12]], dtype=int64)
```

```
In [17]: sklearn.metrics.plot_confusion_matrix(dtree, x_test,y_test)
```

```
Out[17]: <sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x1ef7eeb38e0>
```

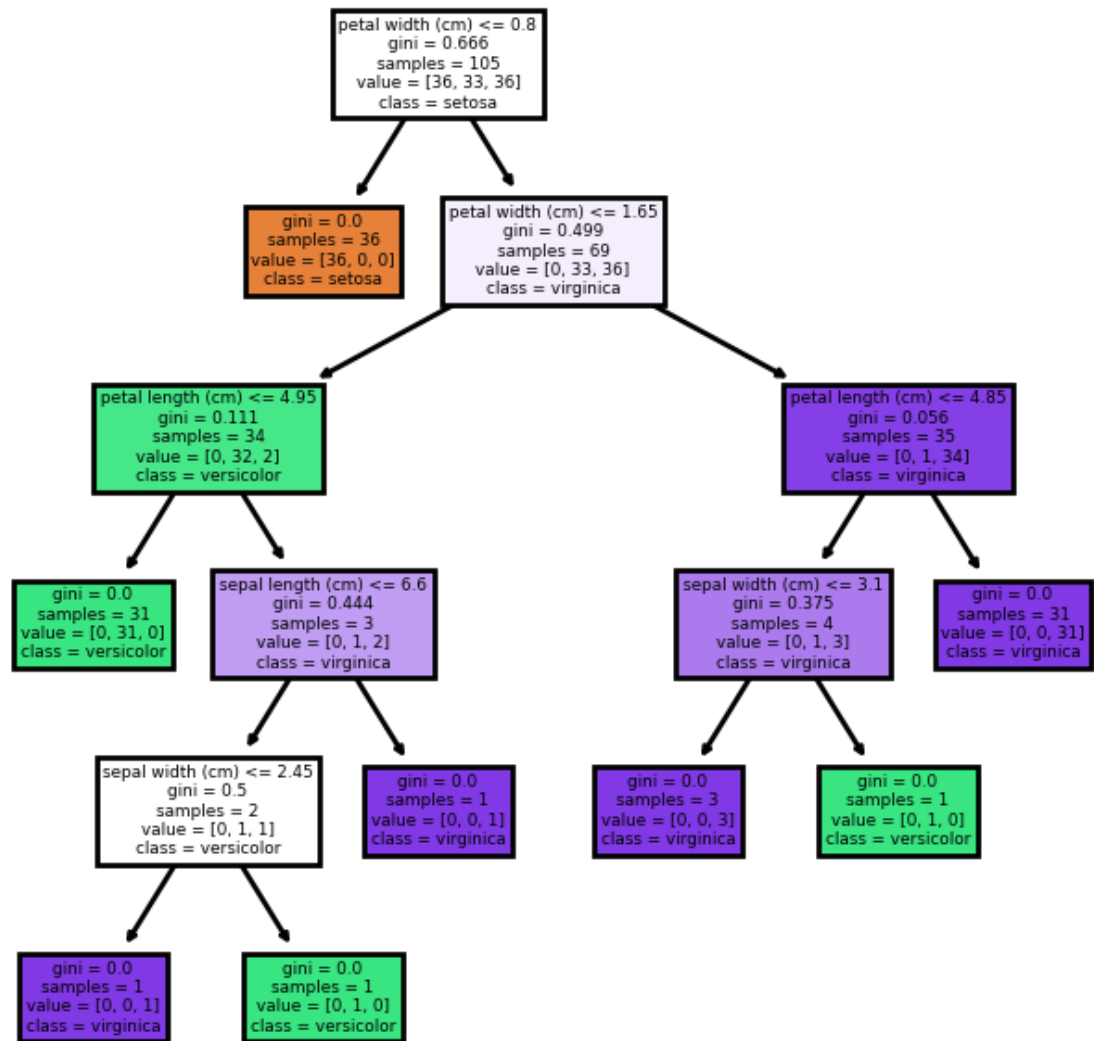


```
In [18]: pip install graphviz
```

Requirement already satisfied: graphviz in c:\users\ashmitabhattacharjee\anaconda3\lib\site-packages (0.17)

Note: you may need to restart the kernel to use updated packages.

```
In [21]: fn=['sepal length (cm)','sepal width (cm)','petal length (cm)','petal width (cm)']
cn=['setosa', 'versicolor', 'virginica']
fig, axes = plt.subplots(nrows = 1,ncols = 1,figsize = (4,4), dpi=200)
tree.plot_tree(dtree,feature_names = fn, class_names=cn,filled = True);
```



In []: