In [1]:

```
import pandas as pd
import numpy as np
from matplotlib import pyplot as plt
import seaborn as sns
from sklearn.preprocessing import LabelEncoder, OneHotEncoder
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import confusion_matrix,classification_report
```

In [2]:

```
#Reading the file
gd=pd.read_csv(r"C:\Users\Agnel Sharon Jerald\OneDrive\Desktop\Machine learning\Iris.csv")
gd
```

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
...	...	...	...	...	...	...
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

In [3]:

```
gd.shape
```

Out[3]:

```
(150, 6)
```

In [4]:

```
gd.dtypes
```

Out[4]:

```
Id                 int64
SepalLengthCm    float64
SepalWidthCm     float64
PetalLengthCm    float64
PetalWidthCm     float64
Species           object
dtype: object
```

In [5]:

```
gd.isna().sum()
```

Out[5]:

```
Id               0
SepalLengthCm    0
SepalWidthCm     0
PetalLengthCm    0
PetalWidthCm     0
Species          0
dtype: int64
```

In [6]:

```
x=gd.iloc[:,:-1]
y=gd.iloc[:,-1]
```

In [7]:

```
ohe=OneHotEncoder(drop='first',sparse_output=False)
x_trns = ohe.fit_transform(x)
x_trns
```

Out[7]:

```
array([[0., 0., 0., ..., 0., 0., 0.],
       [1., 0., 0., ..., 0., 0., 0.],
       [0., 1., 0., ..., 0., 0., 0.],
       ...,
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 1., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.]])
```

In [8]:

```
x_trns=pd.DataFrame(x_trns,columns=ohe.get_feature_names_out(x.columns))
print(x_trns)
```

```
     Id_2  Id_3  Id_4  Id_5  Id_6  Id_7  Id_8  Id_9  Id_10  Id_11  ...  \
0     0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0    0.0    0.0  ...   
1     1.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0    0.0    0.0  ...   
2     0.0   1.0   0.0   0.0   0.0   0.0   0.0   0.0    0.0    0.0  ...   
3     0.0   0.0   1.0   0.0   0.0   0.0   0.0   0.0    0.0    0.0  ...   
4     0.0   0.0   0.0   1.0   0.0   0.0   0.0   0.0    0.0    0.0  ...   
..    ...   ...   ...   ...   ...   ...   ...   ...    ...    ...  ...   
145   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0    0.0    0.0  ...   
146   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0    0.0    0.0  ...   
147   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0    0.0    0.0  ...   
148   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0    0.0    0.0  ...   
149   0.0   0.0   0.0   0.0   0.0   0.0   0.0   0.0    0.0    0.0  ...   

     PetalWidthCm_1.6  PetalWidthCm_1.7  PetalWidthCm_1.8  PetalWidthCm_1.9  \
0                 0.0               0.0               0.0               0.0   
1                 0.0               0.0               0.0               0.0   
2                 0.0               0.0               0.0               0.0   
3                 0.0               0.0               0.0               0.0   
4                 0.0               0.0               0.0               0.0   
..                ...               ...               ...               ...   
145               0.0               0.0               0.0               0.0   
146               0.0               0.0               0.0               1.0   
147               0.0               0.0               0.0               0.0   
148               0.0               0.0               0.0               0.0   
149               0.0               0.0               1.0               0.0   

     PetalWidthCm_2.0  PetalWidthCm_2.1  PetalWidthCm_2.2  PetalWidthCm_2.3  \
0                 0.0               0.0               0.0               0.0   
1                 0.0               0.0               0.0               0.0   
2                 0.0               0.0               0.0               0.0   
3                 0.0               0.0               0.0               0.0   
4                 0.0               0.0               0.0               0.0   
..                ...               ...               ...               ...   
145               0.0               0.0               0.0               1.0   
146               0.0               0.0               0.0               0.0   
147               1.0               0.0               0.0               0.0   
148               0.0               0.0               0.0               1.0   
149               0.0               0.0               0.0               0.0   

     PetalWidthCm_2.4  PetalWidthCm_2.5  
0                 0.0               0.0  
1                 0.0               0.0  
2                 0.0               0.0  
3                 0.0               0.0  
4                 0.0               0.0  
..                ...               ...  
145               0.0               0.0  
146               0.0               0.0  
147               0.0               0.0  
148               0.0               0.0  
149               0.0               0.0  

[150 rows x 268 columns]
```

In [9]:

```
x_trns.dtypes
```

Out[9]:

```
Id_2                float64
Id_3                float64
Id_4                float64
Id_5                float64
Id_6                float64
                     ...   
PetalWidthCm_2.1    float64
PetalWidthCm_2.2    float64
PetalWidthCm_2.3    float64
PetalWidthCm_2.4    float64
PetalWidthCm_2.5    float64
Length: 268, dtype: object
```

In [10]:

```
le = LabelEncoder()
y_trns=le.fit_transform(y)
y_trns
```

Out[10]:

```
array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
       2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
       2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2])
```

In [11]:

```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x_trns,y_trns,test_size=0.2,random_state=42)
```

In [12]:

```
print(x_train.shape,x_test.shape,y_train.shape,y_test.shape)
```

```
(120, 268) (30, 268) (120,) (30,)
```

In [13]:

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

In [14]:

```
#Model Building
dt= DecisionTreeClassifier()
dt.fit(x_train,y_train)
y_train_pred=dt.predict(x_train)
y_test_pred=dt.predict(x_test)
```

In [15]:

```
#train data
confusion_matrix(y_train,y_train_pred)
```

Out[15]:

```
array([[40,  0,  0],
       [ 0, 41,  0],
       [ 0,  0, 39]])
```

In [16]:

```
#test data
confusion_matrix(y_test,y_test_pred)
```

Out[16]:

```
array([[ 9,  0,  1],
       [ 0,  8,  1],
       [ 0,  0, 11]])
```

In [17]:

```
#classification report
print(classification_report(y_train,y_train_pred))
```

```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        40
           1       1.00      1.00      1.00        41
           2       1.00      1.00      1.00        39

    accuracy                           1.00       120
   macro avg       1.00      1.00      1.00       120
weighted avg       1.00      1.00      1.00       120
```

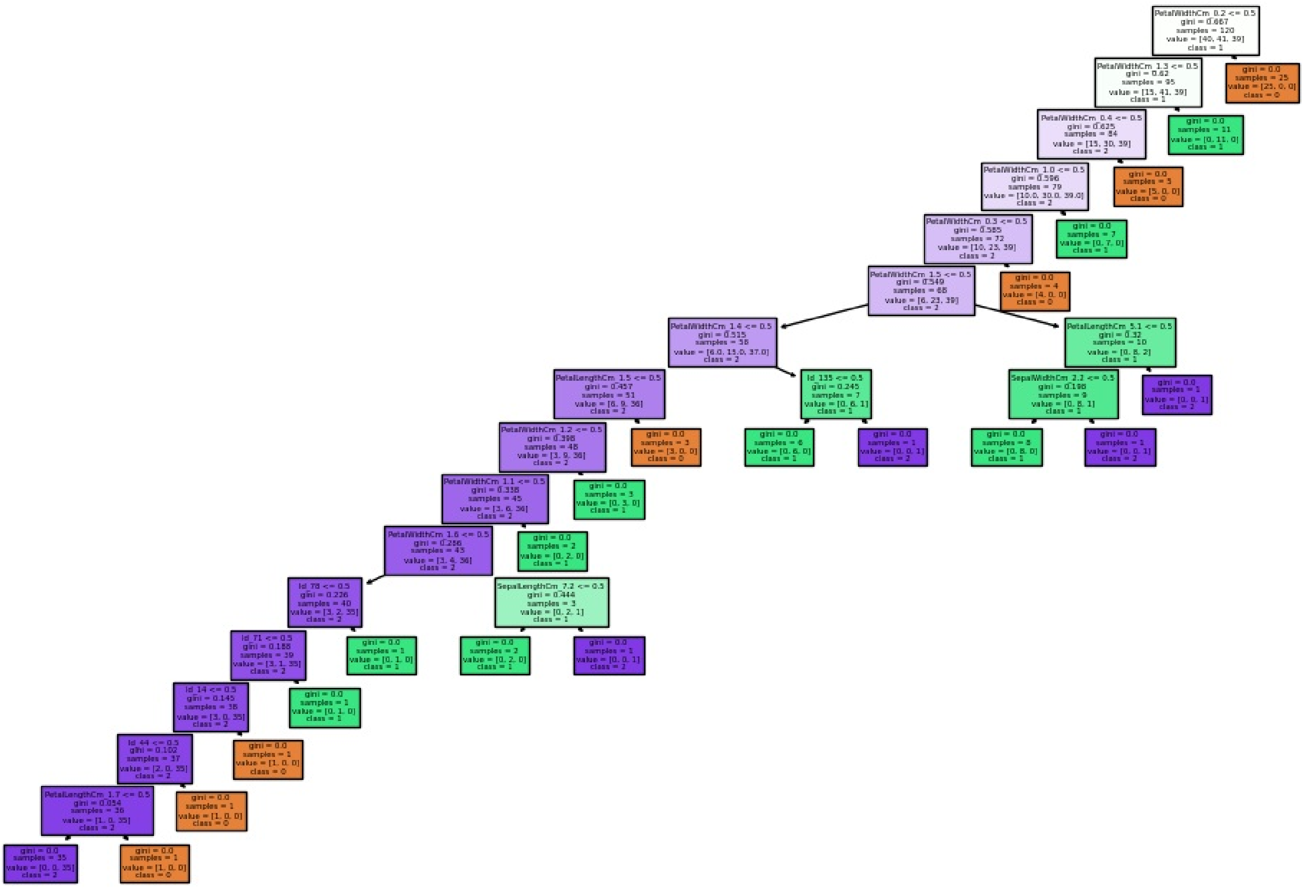
In [18]:

```
print(classification_report(y_test,y_test_pred))
```

```
              precision    recall  f1-score   support

           0       1.00      0.90      0.95        10
           1       1.00      0.89      0.94         9
           2       0.85      1.00      0.92        11

    accuracy                           0.93        30
   macro avg       0.95      0.93      0.94        30
weighted avg       0.94      0.93      0.93        30
```

In [20]:

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

plt.figure(figsize=(12, 8))
tree.plot_tree(
    dt,
    feature_names=x_trns.columns,
    class_names=dt.classes_.astype(str),
    filled=True
)
plt.show()
```


PetalWidthCm 0.2 <= 0.5
gini = 0.667
samples = 120
value = [40, 41, 39]
class = 1
PetalWidthCm 1.3 <= 0.5
gini = 0.62
samples = 95
value = [15, 41, 39]
class = 1
gini = 0.0
samples = 25
value = [25, 0, 0]
class = 0
PetalWidthCm 0.4 <= 0.5
gini = 0.625
samples = 84
value = [15, 30, 39]
class = 2
gini = 0.0
samples = 11
value = [0, 11, 0]
class = 1
PetalWidthCm 1.0 <= 0.5
gini = 0.596
samples = 79
value = [10.0, 30.0, 39.0]
class = 2
gini = 0.0
samples = 5
value = [5, 0, 0]
class = 0
PetalWidthCm 0.3 <= 0.5
gini = 0.585
samples = 72
value = [10, 23, 39]
class = 2
gini = 0.0
samples = 7
value = [0, 7, 0]
class = 1
PetalWidthCm 1.5 <= 0.5
gini = 0.549
samples = 68
value = [6, 23, 39]
class = 2
gini = 0.0
samples = 4
value = [4, 0, 0]
class = 0
PetalWidthCm 1.4 <= 0.5
gini = 0.515
samples = 58
value = [6.0, 15.0, 37.0]
class = 2
PetalLengthCm 5.1 <= 0.5
gini = 0.32
samples = 10
value = [0, 8, 2]
class = 1
PetalLengthCm 1.5 <= 0.5
gini = 0.457
samples = 51
value = [6, 9, 36]
class = 2
Id 135 <= 0.5
gini = 0.245
samples = 7
value = [0, 6, 1]
class = 1
SepalWidthCm 2.2 <= 0.5
gini = 0.198
samples = 9
value = [0, 8, 1]
class = 1
gini = 0.0
samples = 1
value = [0, 0, 1]
class = 2
PetalWidthCm 1.2 <= 0.5
gini = 0.398
samples = 48
value = [3, 9, 36]
class = 2
gini = 0.0
samples = 3
value = [3, 0, 0]
class = 0
gini = 0.0
samples = 6
value = [0, 6, 0]
class = 1
gini = 0.0
samples = 1
value = [0, 0, 1]
class = 2
gini = 0.0
samples = 8
value = [0, 8, 0]
class = 1
gini = 0.0
samples = 1
value = [0, 0, 1]
class = 2
PetalWidthCm 1.1 <= 0.5
gini = 0.338
samples = 45
value = [3, 6, 36]
class = 2
gini = 0.0
samples = 3
value = [0, 3, 0]
class = 1
PetalWidthCm 1.6 <= 0.5
gini = 0.286
samples = 43
value = [3, 4, 36]
class = 2
gini = 0.0
samples = 2
value = [0, 2, 0]
class = 1
Id 78 <= 0.5
gini = 0.226
samples = 40
value = [3, 2, 35]
class = 2
SepalLengthCm 7.2 <= 0.5
gini = 0.444
samples = 3
value = [0, 2, 1]
class = 1
Id 71 <= 0.5
gini = 0.188
samples = 39
value = [3, 1, 35]
class = 2
gini = 0.0
samples = 1
value = [0, 1, 0]
class = 1
gini = 0.0
samples = 2
value = [0, 2, 0]
class = 1
gini = 0.0
samples = 1
value = [0, 0, 1]
class = 2
Id 14 <= 0.5
gini = 0.145
samples = 38
value = [3, 0, 35]
class = 2
gini = 0.0
samples = 1
value = [0, 1, 0]
class = 1
Id 44 <= 0.5
gini = 0.102
samples = 37
value = [2, 0, 35]
class = 2
gini = 0.0
samples = 1
value = [1, 0, 0]
class = 0
PetalLengthCm 1.7 <= 0.5
gini = 0.054
samples = 36
value = [1, 0, 35]
class = 2
gini = 0.0
samples = 1
value = [1, 0, 0]
class = 0
gini = 0.0
samples = 35
value = [0, 0, 35]
class = 2
gini = 0.0
samples = 1
value = [1, 0, 0]
class = 0