importing libraries

In [1]:
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.tree import DecisionTreeClassifier, plot_tree
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import MinMaxScaler, LabelEncoder
from sklearn.metrics import accuracy_score, confusion_matrix, precision_score
import math
import numpy as np
```

reading file

In [2]:
```
iris = pd.read_csv("iris.csv")
```

seperating label and features

In [3]:
```
x = iris.drop("species", axis = 1)
y = iris["species"]
```

training and testing

In [4]:
```
clasi = DecisionTreeClassifier()
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.2, ranc
clasi.fit(x_train, y_train)
```

Out[4]:
```
DecisionTreeClassifier
DecisionTreeClassifier()
```

getting accuracy

In [5]:
```
accuracy = clasi.score(x_test, y_test)
print("accuracy score", accuracy)
```

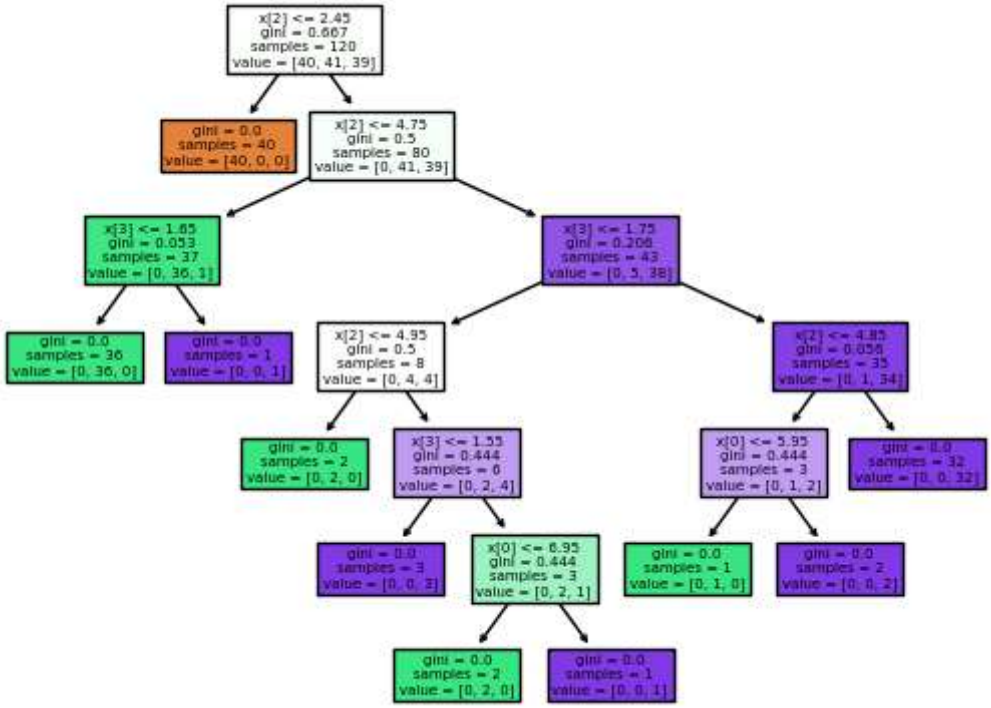
```
accuracy score 1.0
```

making predicts

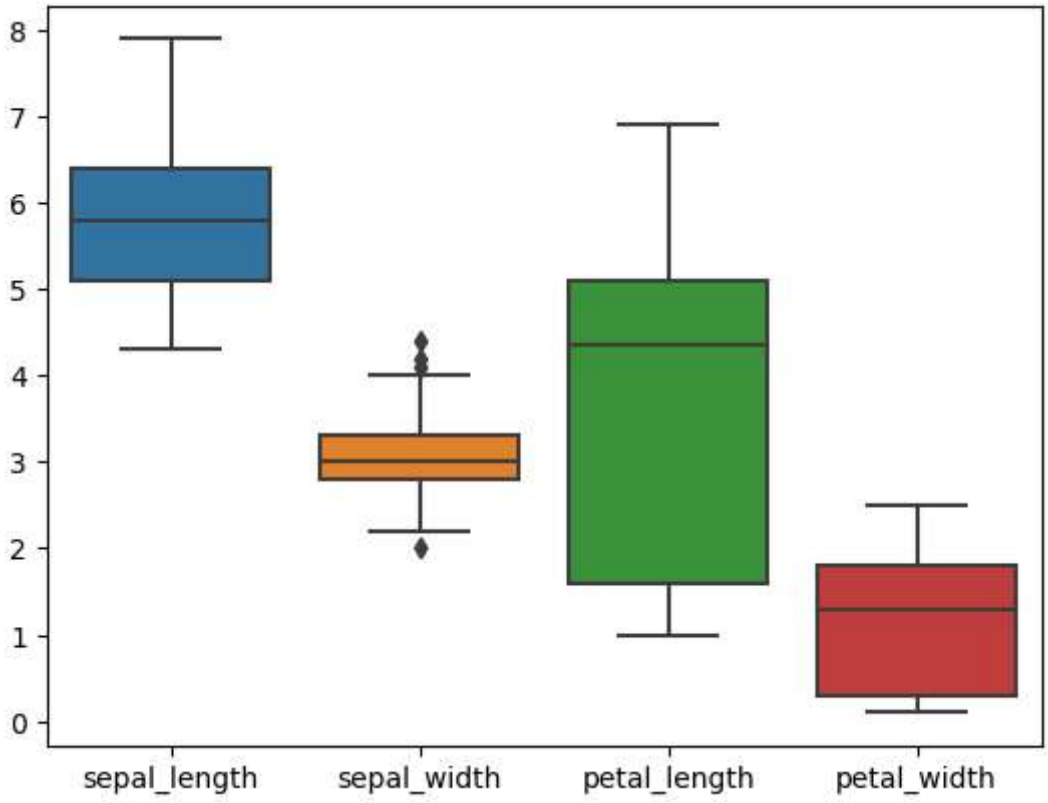
In [6]:
```
y_preds = clasi.predict(x_test)
```

decision tree

In [7]:
```
plot_tree(clasi, filled = True);
```



In [8]:
```
sns.boxplot(data = iris);
```



In [9]:
```
tot_species = len(y)
class_counts = y.value_counts()
entropy = 0
for count in class_counts:
    prob = count / tot_species
    entropy = entropy - prob *  math.log2(prob)
print(f"Entropy: {entropy}")
```

```
Entropy: 1.584962500721156
```

In [11]:
```
import numpy as np

def gini_index(labels):
    classes, count = np.unique(labels, return_counts=True)

    prob = count / len(labels)
    gini = 1 - np.sum(prob ** 2)
    return gini



gini = gini_index(y)
print("Gini Index:", gini)
```

```
  Cell In[11], line 13
    print("Gini Index:", gini)
                        ^
SyntaxError: invalid non-printable character U+00A0
```

In [ ]:
```

```