Module 3 Decision Tree Classifier

In [1]:

```
import pandas as p
import matplotlib.pyplot as plt
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import confusion_matrix, classification_report, accuracy_score,
```

In [2]:

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

In [3]:

```
data = p.read_csv('crop.csv')
```

In [4]:

```
df=data.dropna()
```

In [5]:

```
df.columns
```

Out[5]:

```
Index(['nitrogen', 'phosphorus', 'potassium', 'temperature', 'humidity', 'ph',
       'rainfall', 'label'],
      dtype='object')
```

In [6]:

```
var_mod =['label']
le = LabelEncoder()
for i in var_mod:
    df[i] = le.fit_transform(df[i]).astype(int)
```

In [7]:

```
inputs = df.drop(labels='label', axis=1) #takes all features except label as input
output = df.loc[:,'label'] #Returns label(crop type) as output
```

In [8]:

```
X_train, X_test, y_train, y_test = train_test_split(inputs, output, test_size=0.3, r
print("Number of Training Datasets: ", len(X_train))
print("Number of Testing Dataset: ", len(X_test))
print("Total Number of Datasets: ", len(X_train)+len(X_test))
```

```
Number of Training Datasets:  1540
Number of Testing Dataset:  660
Total Number of Datasets:  2200
```

In [9]:

```
#Model Training
dt = DecisionTreeClassifier()
dt.fit(X_train,y_train)
predicted_dt = dt.predict(X_test)
```

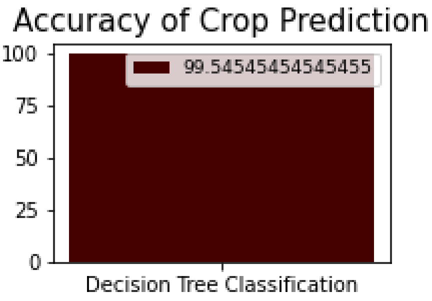
In [10]:

```
#Getting Accuracy
accuracy = accuracy_score(y_test,predicted_dt) # accuracy: (tp + tn) / (p + n)
print('Accuracy of Decision Tree Classifier is: ',accuracy*100)
```

```
Accuracy of Decision Tree Classifier is:  99.54545454545455
```

In [11]:

```
DT=accuracy.mean() *100
def graph():
    data=[DT]
    alg="Decision Tree Classification"
    plt.figure(figsize=(3,2))
    b=plt.bar(alg,data,color=("#450000"))
    plt.title("Accuracy of Crop Prediction",fontsize=15)
    plt.legend(b,data,fontsize=9)
graph()
```



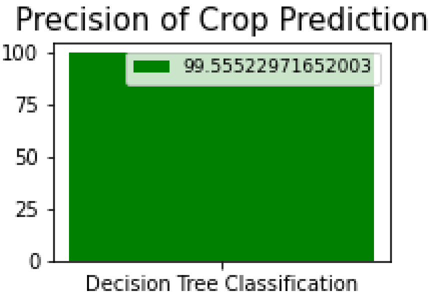
In [12]:

```
#Getting Precision
precision = precision_score(y_test,predicted_dt,average='weighted') # precision tp /
#weighted average takes mean of all classes' scores
print('Precision of Decision Tree Classifier is: ',precision*100)
```

```
Precision of Decision Tree Classifier is:  99.55522971652003
```

In [13]:

```
DT=precision.mean() *100
def graph():
    data=[DT]
    alg="Decision Tree Classification"
    plt.figure(figsize=(3,2))
    b=plt.bar(alg,data,color=("green"))
    plt.title("Precision of Crop Prediction",fontsize=15)
    plt.legend(b,data,fontsize=9)
graph()
```



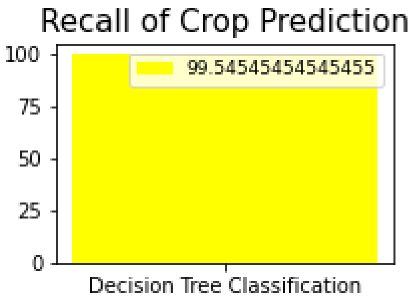
In [14]:

```
#Getting Recall
recall = recall_score(y_test,predicted_dt,average='weighted') # recall: tp / (tp + f
print('Recall of Decision Tree Classifier is: ',recall*100)
```

```
Recall of Decision Tree Classifier is:  99.54545454545455
```

In [15]:

```
DT=recall.mean() *100
def graph():
    data=[DT]
    alg="Decision Tree Classification"
    plt.figure(figsize=(3,2))
    b=plt.bar(alg,data,color=("yellow"))
    plt.title("Recall of Crop Prediction",fontsize=15)
    plt.legend(b,data,fontsize=9)
graph()
```



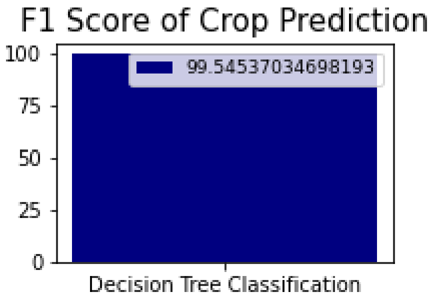
In [16]:

```
#Getting F1 Score
f1 = f1_score(y_test,predicted_dt,average='weighted') # f1: 2 tp / (2 tp + fp + fn)
print('F1 Score of Decision Tree Classifier is: ',f1*100)
```

```
F1 Score of Decision Tree Classifier is:  99.54537034698193
```

In [17]:

```
DT=f1.mean() *100
def graph():
    data=[DT]
    alg="Decision Tree Classification"
    plt.figure(figsize=(3,2))
    b=plt.bar(alg,data,color=("#000080"))
    plt.title("F1 Score of Crop Prediction",fontsize=15)
    plt.legend(b,data,fontsize=9)
graph()
```



In [18]:

```
#Classification Report
cr = classification_report(y_test,predicted_dt)
print('Classification report\n',cr)
```

```
Classification report
               precision    recall  f1-score   support

           0       1.00      1.00      1.00        30
           1       1.00      1.00      1.00        30
           2       0.97      1.00      0.98        30
           3       1.00      1.00      1.00        30
           4       1.00      1.00      1.00        30
           5       1.00      0.97      0.98        30
           6       0.97      1.00      0.98        30
           7       1.00      1.00      1.00        30
           8       1.00      1.00      1.00        30
           9       1.00      1.00      1.00        30
          10       1.00      1.00      1.00        30
          11       0.97      0.97      0.97        30
          12       1.00      1.00      1.00        30
          13       1.00      0.97      0.98        30
          14       1.00      1.00      1.00        30
          15       1.00      1.00      1.00        30
          16       1.00      1.00      1.00        30
          17       1.00      1.00      1.00        30
          18       1.00      1.00      1.00        30
          19       1.00      1.00      1.00        30
          20       1.00      1.00      1.00        30
          21       1.00      1.00      1.00        30

    accuracy                           1.00       660
   macro avg       1.00      1.00      1.00       660
weighted avg       1.00      1.00      1.00       660
```

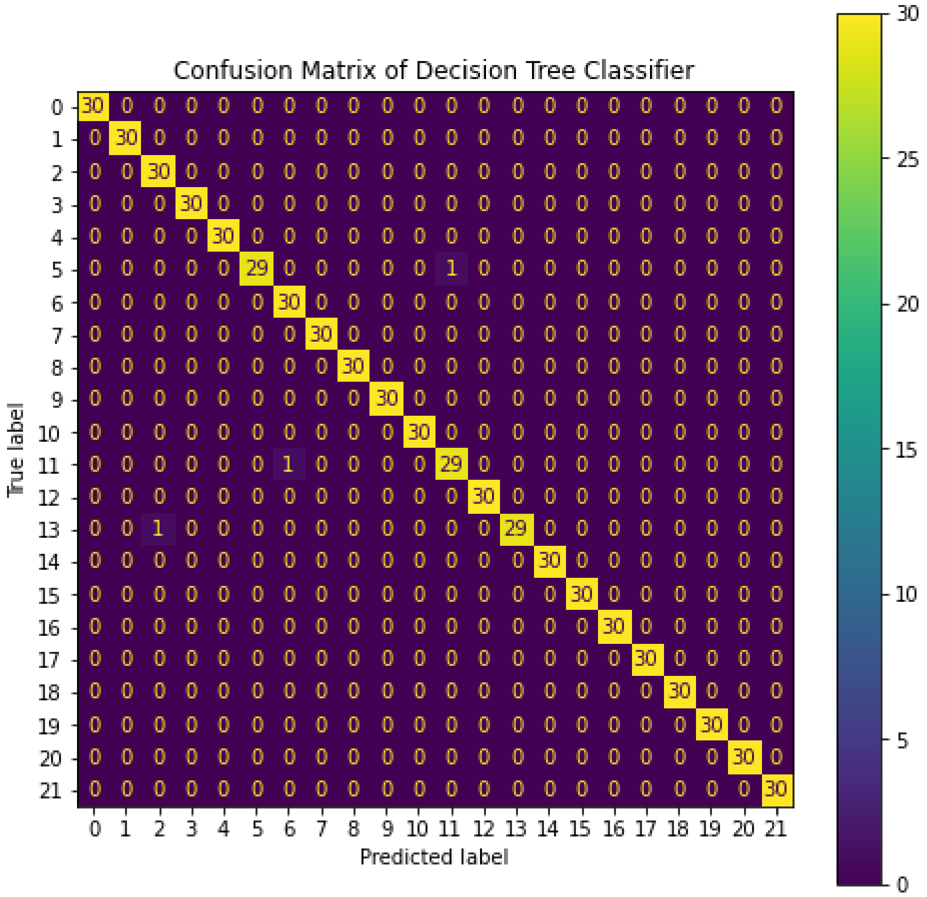
In [19]:

```
#Confusion Matrix
cm = confusion_matrix(y_test,predicted_dt)
print('Confusion matrix\n',cm)
```

```
Confusion matrix
 [[30  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0 30  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0 30  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0 30  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0 30  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0 29  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0 30  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0 30  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0 30  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0 30  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0 30  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  1  0  0  0  0 29  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0 30  0  0  0  0  0  0  0  0  0]
 [ 0  0  1  0  0  0  0  0  0  0  0  0  0 29  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0 30  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 30  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 30  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 30  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 30  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 30  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 30  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 30]]
```

In [20]:

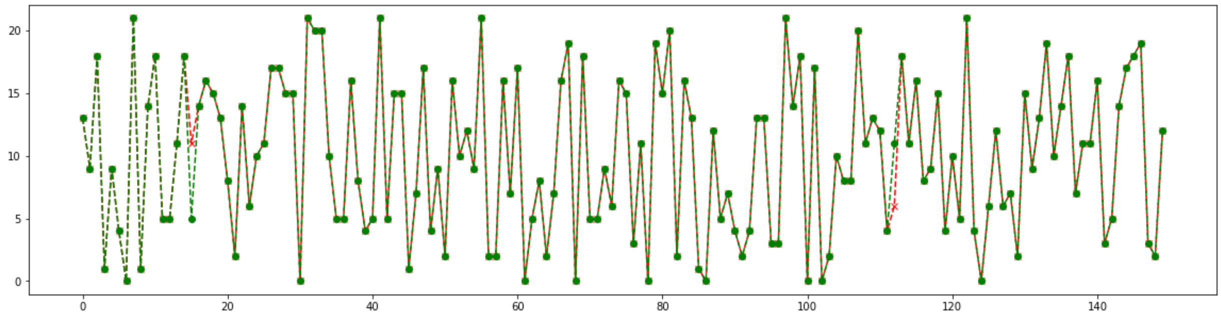
```
import matplotlib.pyplot as plt
fig, ax = plt.subplots(figsize=(8,8))
plot_confusion_matrix(dt, X_test, y_test, ax=ax)
plt.title('Confusion Matrix of Decision Tree Classifier')
plt.show()
```



	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
0	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	29	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	1	0	0	0	0	29	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0
13	0	0	1	0	0	0	0	0	0	0	0	0	0	29	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30

In [21]:

```
DF = p.DataFrame()
DF["y_test"] = y_test
DF["predicted"] = predicted_dt
DF.reset_index(inplace=True)
plt.figure(figsize=(20, 5))
plt.plot(DF["predicted"][:150], marker='x', linestyle='dashed', color='red')
plt.plot(DF["y_test"][:150],  marker='o', linestyle='dashed', color='green')
plt.show()
```



In [22]:

```
#Saving Model
import joblib
joblib.dump(dt,'dt.pkl')
```

Out[22]:

```
['dt.pkl']
```