

## Exercise 2.2

### 1.1 Problem Statement:

To Implement ID3 (information gain) algorithm for decision tree learning for transforming continuous variables into discrete variables.

### 1.2 Description of Machine learning Algorithm:

ID3 (Iterative Dichotomiser 3) is named such because the algorithm iteratively (repeatedly) dichotomizes(divides) features into two or more groups at each step.

Invented by Ross Quinlan ID3 algorithm builds decision trees using a top-down greedy search approach through the space of possible branches with no backtracking. Classification algorithm that follows a greedy approach by selecting a best attribute that yields maximum Information Gain(IG) or minimum Entropy(H). Typically used in Machine Learning and Natural Language Processing domains.

Some major benefits of ID3 are: **Understandable prediction rules are created from the training data.** Builds a short tree in relatively small time. It only needs to test enough attributes until all data is classified.

### 1.3 Description of the Dataset:

Title of the dataset: Iris Plants Database

The Iris Dataset contains information of three species of Iris flowers (Iris setosa, Iris virginica and Iris versicolor). The data set contains 3 classes of 50 instances each, where each class refers to a type of iris plant. One class is linearly separable from the other 2; the latter are NOT linearly separable from each other.

Data Set Characteristics: Multivariate

Area: Life Sciences

Number of samples (or instances) in the dataset: 150

Number of attributes (or features): 05 Attribute

Information:

1. sepal length in cm
2. sepal width in cm
3. petal length in cm
4. petal width in cm
5. class:
  - Iris Setosa
  - Iris Versicolour
  - Iris Virginica
6. Number of samples of each species of iris flowers:
  - Class Distribution: 33.3% for each of 3 classes.
  - 50 (Setosa), 50 (Versicolor), 50 (Virginica)
7. Predicted attribute: class of iris plant.
8. Missing Attribute Values: None

Feature Name	Units of measurement	Datatype	Description
sepal length	Centimeters	Real (Numerical)	Length of Iris flower's sepal
sepal width length	Centimeters	Real (Numerical)	Width of Iris flower's sepal
petal length	Centimeters	Real (Numerical)	Length of Iris flower's petal
petal width length	Centimeters	Real (Numerical)	Width of Iris flower's petal
variety	Variety of species [Setosa, Virginica, Versicolor]	Object (Categorical)	Variety of the species of Iris flower

## 1.4 Data Preprocessing and Exploratory Data Analysis (EDA):

Data preprocessing is the process of transforming raw data into an understandable format. It is also an important step in data mining as we cannot work with raw data. The quality of the data should be checked before applying machine learning or data mining algorithms.

Major Tasks in Data Preprocessing:

1. Data cleaning
2. Data integration
3. Data reduction
4. Data transformation

Exploratory data analysis (EDA) is a technique that data professionals can use to understand a dataset before they start to model it. Some people refer to EDA as data exploration. The goal of conducting EDA is to determine the characteristics of the dataset. Conducting EDA can help data analysts make predictions and assumptions about data. Often, EDA involves data visualization, including creating graphs like histograms, scatter plots and box plots.

Major Tasks in EDA:

1. Observe your dataset
2. Find any missing values
3. Categorize your values
4. Find the shape of your dataset
5. Identify relationships in your dataset
6. Locate any outliers in your dataset

## 1.5 Machine Learning Package Used for Model building:

The scikit-learn (formerly scikits.learn and also known as sklearn) is a free software machine learning library for the Python programming language. It provides simple and efficient tools for predictive data analysis. It features various classification, regression and clustering algorithms including support-vector machines, random forests, gradient boosting, k-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy.

The `sklearn.model_selection.train_test_split()` method splits arrays or matrices into random train and test subsets. The parameters for the method are as follows:

```
sklearn.model_selection.train_test_split(*arrays, test_size=None, train_size=None,
random_state=None, shuffle=True, stratify=None)
```

It returns lists containing train-test split of inputs.

## 1.6 Implementation:

```
from google.colab import files
uploaded=files.upload()
from sklearn.datasets import load_iris
iris = load_iris()
X = iris.data
y = iris.target
feature_names = iris.feature_names
target_names = iris.target_names
print("Feature names:", feature_names)
print("Target names:", target_names)
print("\nFirst 10 rows of X:\n", X[:10])
Feature names: ['sepal length (cm)', 'sepal width (cm)', 'petal length (cm)',
'petal width (cm)']
Target names: ['setosa' 'versicolor' 'virginica']

First 10 rows of X:
[[5.1 3.5 1.4 0.2]
 [4.9 3.  1.4 0.2]
 [4.7 3.2 1.3 0.2]
 [4.6 3.1 1.5 0.2]
 [5.  3.6 1.4 0.2]
 [5.4 3.9 1.7 0.4]
 [4.6 3.4 1.4 0.3]
 [5.  3.4 1.5 0.2]
 [4.4 2.9 1.4 0.2]
 [4.9 3.1 1.5 0.1]]

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size = 0.3, random_state = 1
)

print(X_train.shape)
print(X_test.shape)

print(y_train.shape)
print(y_test.shape)
(105, 4)
(45, 4)
(105,)
(45,)

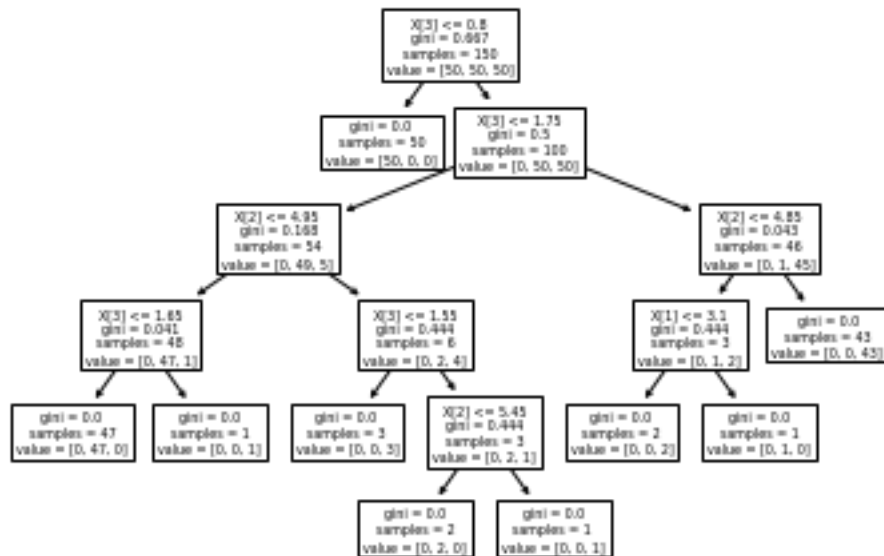
from sklearn import tree
clf = tree.DecisionTreeClassifier()
clf = clf.fit(X, y)

tree.plot_tree(clf)
_ [Text(0.5, 0.9166666666666666, 'X[3] <= 0.8\ngini = 0.667\nsamples = 150\nvalue =
[50, 50, 50]'),
  Text(0.4230769230769231, 0.75, 'gini = 0.0\nsamples = 50\nvalue = [50, 0, 0]'),
```

```

Text(0.5769230769230769, 0.75, 'X[3] <= 1.75\n'gini = 0.5\n'samples = 100\n'value =
[0, 50, 50]'),
Text(0.3076923076923077, 0.5833333333333334, 'X[2] <= 4.95\n'gini = 0.168\n'samples
= 54\n'value = [0, 49, 5]'),
Text(0.15384615384615385, 0.4166666666666667, 'X[3] <= 1.65\n'gini =
0.041\n'samples = 48\n'value = [0, 47, 1]'),
Text(0.07692307692307693, 0.25, 'gini = 0.0\n'samples = 47\n'value = [0, 47, 0]'),
Text(0.23076923076923078, 0.25, 'gini = 0.0\n'samples = 1\n'value = [0, 0, 1]'),
Text(0.46153846153846156, 0.4166666666666667, 'X[3] <= 1.55\n'gini =
0.444\n'samples = 6\n'value = [0, 2, 4]'),
Text(0.38461538461538464, 0.25, 'gini = 0.0\n'samples = 3\n'value = [0, 0, 3]'),
Text(0.5384615384615384, 0.25, 'X[2] <= 5.45\n'gini = 0.444\n'samples = 3\n'value =
[0, 2, 1]'),
Text(0.46153846153846156, 0.08333333333333333, 'gini = 0.0\n'samples = 2\n'value =
[0, 2, 0]'),
Text(0.6153846153846154, 0.08333333333333333, 'gini = 0.0\n'samples = 1\n'value =
[0, 0, 1]'),
Text(0.8461538461538461, 0.5833333333333334, 'X[2] <= 4.85\n'gini = 0.043\n'samples
= 46\n'value = [0, 1, 45]'),
Text(0.7692307692307693, 0.4166666666666667, 'X[1] <= 3.1\n'gini = 0.444\n'samples
= 3\n'value = [0, 1, 2]'),
Text(0.6923076923076923, 0.25, 'gini = 0.0\n'samples = 2\n'value = [0, 0, 2]'),
Text(0.8461538461538461, 0.25, 'gini = 0.0\n'samples = 1\n'value = [0, 1, 0]'),
Text(0.9230769230769231, 0.4166666666666667, 'gini = 0.0\n'samples = 43\n'value =
[0, 0, 43]')

```

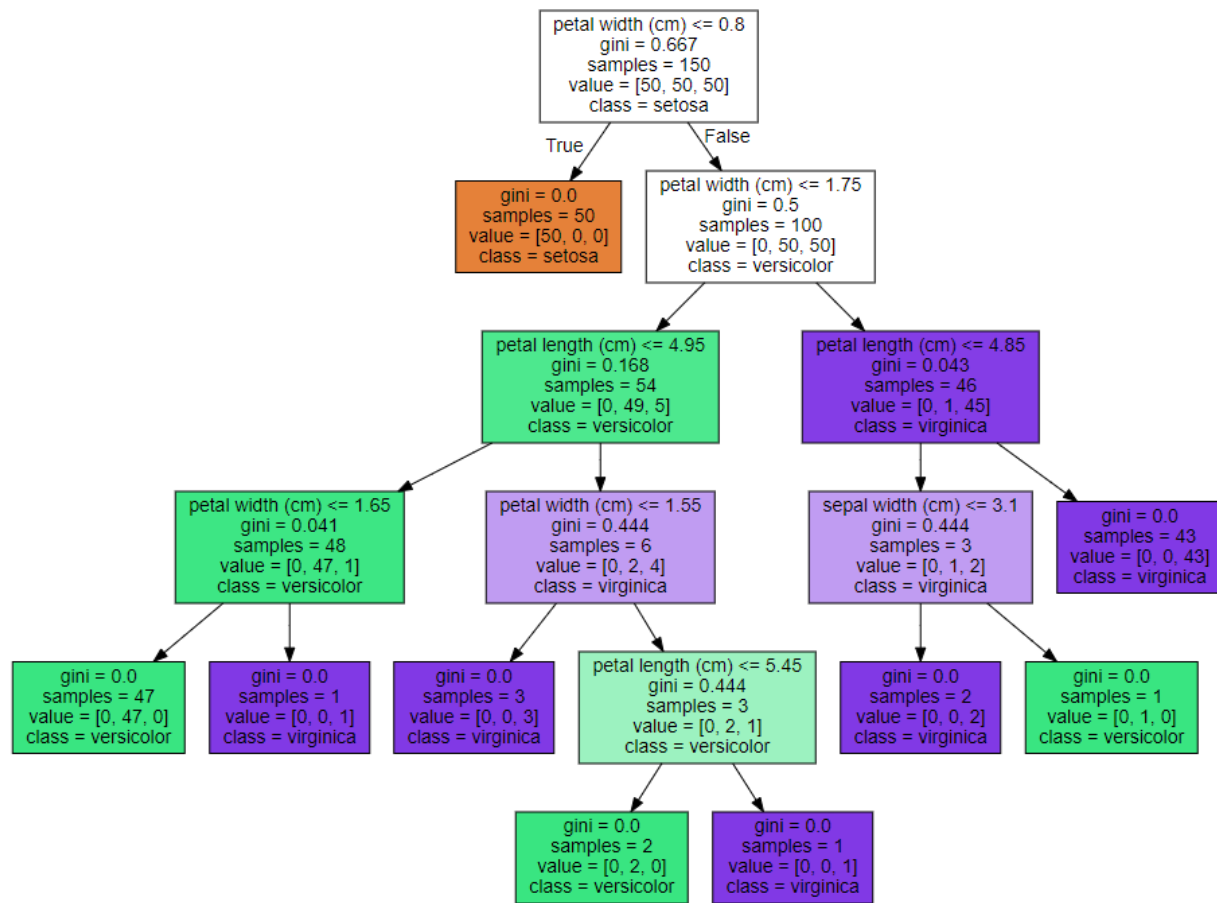


## Graphical Representation:

```

#graphviz
import graphviz
dot_data = tree.export_graphviz(clf, out_file=None,
                                feature_names=iris.feature_names,
                                class_names=iris.target_names,
                                filled=True)
graph = graphviz.Source(dot_data, format="png")
graph

```



```

from sklearn import metrics
#Predict the response for test dataset
y_pred = clf.predict(X_test)
#Model Accuracy
print("Accuracy:",metrics.accuracy_score(y_test, y_pred))
Accuracy: 1.0

```

Implement ID3 (information gain) algorithm for decision tree learning for transforming continuous variables into discrete variables.

```

import itertools
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from pandas import DataFrame, Series
from IPython.display import Image
try:
    from StringIO import StringIO
except ImportError:
    from io import StringIO
import pydotplus
from sklearn import preprocessing
from sklearn import tree
%matplotlib inline

```

### plotting decision tree using method

```
def plot_decision_tree(clf, feature_name, target_name):
    dot_data = StringIO()
    tree.export_graphviz(clf, out_file=dot_data,
                        feature_names=feature_name,
                        class_names=target_name,
                        filled=True, rounded=True,
                        special_characters=True)
    graph = pydotplus.graph_from_dot_data(dot_data.getvalue())
    return Image(graph.create_png())
df=pd.read_csv("Iris.csv")
df
```

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

```
df['Species'].unique()
df['Species']=df['Species'].replace(['Iris-setosa','Iris-versicolor','Iris-
virginica'],[0,1,2])
df['Species']
0      0
1      0
2      0
3      0
4      0
..
145    2
146    2
147    2
148    2
149    2
Name: Species, Length: 150, dtype: int64
```

```
X_train = df[['SepalLengthCm']]
Y_train = df.Species
```

## Method to Calculate Entropy

```
def entropy(target_col):
    elements, counts = np.unique(target_col, return_counts=True)
    for (e, c) in zip(elements, counts):
        print("Node has {} elements of Class {}".format(c, e))
    entropy = np.sum(
        [
            (-
counts[i] / np.sum(counts)) * np.log2(counts[i] / np.sum(counts))
            for i in range(len(elements))
        ]
    )
    return entropy
```

```
entropy_before_split = entropy(df.Species)
print ("Initial Entropy = {}".format(entropy_before_split))
Node has 50 elements of Class 0
Node has 50 elements of Class 1
Node has 50 elements of Class 2
Initial Entropy = 1.584962500721156
```

```
a=df['SepalLengthCm'].to_numpy()
a=np.unique(a)
a=np.sort(a)
a
array([4.3, 4.4, 4.5, 4.6, 4.7, 4.8, 4.9, 5. , 5.1, 5.2, 5.3, 5.4, 5.5,
       5.6, 5.7, 5.8, 5.9, 6. , 6.1, 6.2, 6.3, 6.4, 6.5, 6.6, 6.7, 6.8,
       6.9, 7. , 7.1, 7.2, 7.3, 7.4, 7.6, 7.7, 7.9])
```

```
l=[]
for i in range(len(a)-1):
    k=(a[i]+a[i+1])/2
    l.append(k)
split_values=np.array(l)
```

```
left_node_data = df.loc[df.SepalLengthCm < split_values[0]]
right_node_data = df.loc[df.SepalLengthCm > split_values[0]]
left_node_data
Id   SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm  Species
1314    4.3           3.0           1.1           0.1           0
right_node_data
```

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

149 rows × 6 columns

## Method to Calculate Information Gain

```
def calc_info_gain_numeric(df, left, right, entropy_before_split, split_values
):
    print("Split Value = {} \n".format(split_values))
    total_elements = df.shape[0]
    print("Left Node")
    ent_left = entropy(left)
    print("Entropy of Left Node {} \n".format(ent_left))
    print("Right Node")
    ent_right = entropy(right)
    print("Entropy of Right Node {} \n".format(ent_right))
    weighted_entropy = ((left.shape[0] / total_elements) * ent_left) + (
        (right.shape[0] / total_elements) * ent_right
    )
    info_gain = entropy_before_split - weighted_entropy
    print("Info gain at split {} is {} \n".format(split_values, info_gain))
    return info_gain

calc_info_gain_numeric(
    df, left_node_data.Species, right_node_data.Species, entropy_before_split,
    split_values[0]
)
```



Split Value = 4.35

Left Node

Node has 1 elements of Class 0

Entropy of Left Node 0.0

Right Node

Node has 49 elements of Class 0

Node has 50 elements of Class 1

Node has 50 elements of Class 2

Entropy of Right Node 1.5848973705351974

Info gain at split 4.35 is 0.01063111265619332

0.01063111265619332

```
info_gains = []
for sp_value in split_values:
    left_data = df.loc[df.SepalLengthCm < sp_value]
    right_data = df.loc[df.SepalLengthCm > sp_value]
    info_gains.append(
        calc_info_gain_numeric(
            df, left_data.Species, right_data.Species, entropy_before_split, s
            p_value
        )
    )
```

Split Value = 4.35

Left Node

Node has 1 elements of Class 0

Entropy of Left Node 0.0

Right Node

Node has 49 elements of Class 0

Node has 50 elements of Class 1

Node has 50 elements of Class 2

Entropy of Right Node 1.5848973705351974

Info gain at split 4.35 is 0.01063111265619332

Split Value = 4.45

Left Node

Node has 4 elements of Class 0

Entropy of Left Node 0.0

Right Node

Node has 46 elements of Class 0

Node has 50 elements of Class 1

Node has 50 elements of Class 2

Entropy of Right Node 1.5838692920574484

Info gain at split 4.45 is 0.04332972311857297

Split Value = 4.55

Left Node

Node has 5 elements of Class 0

Entropy of Left Node 0.0

Right Node

Node has 45 elements of Class 0

Node has 50 elements of Class 1

Node has 50 elements of Class 2

Entropy of Right Node 1.5832262740679837

Info gain at split 4.55 is 0.054510435788771794

Split Value = 4.65

Left Node

Node has 9 elements of Class 0

Entropy of Left Node 0.0

Right Node

Node has 41 elements of Class 0

Node has 50 elements of Class 1

Node has 50 elements of Class 2

Entropy of Right Node 1.5789467341944396

Info gain at split 4.65 is 0.10075257057838294

Split Value = 4.75

Left Node

Node has 11 elements of Class 0

Entropy of Left Node 0.0

Right Node

Node has 39 elements of Class 0

Node has 50 elements of Class 1

Node has 50 elements of Class 2

Entropy of Right Node 1.5756582992479666

Info gain at split 4.75 is 0.12485247675137368

Split Value = 4.85

Left Node

Node has 16 elements of Class 0

Entropy of Left Node 0.0

Right Node

Node has 34 elements of Class 0

Node has 50 elements of Class 1

Node has 50 elements of Class 2

Entropy of Right Node 1.5634074323981904

Info gain at split 4.85 is 0.18831852777877267

Split Value = 4.95

Left Node

Node has 20 elements of Class 0

Node has 1 elements of Class 1

Node has 1 elements of Class 2

Entropy of Left Node 0.5304060778306042

Right Node

Node has 30 elements of Class 0

Node has 49 elements of Class 1

Node has 49 elements of Class 2

Entropy of Right Node 1.5511852922535474

Info gain at split 4.95 is 0.18349149324964031

Split Value = 5.05

Left Node

Node has 28 elements of Class 0

Node has 3 elements of Class 1

Node has 1 elements of Class 2

Entropy of Left Node 0.6449742087569881

Right Node

Node has 22 elements of Class 0

Node has 47 elements of Class 1

Node has 49 elements of Class 2

Entropy of Right Node 1.5072705578320558

Info gain at split 5.05 is 0.2616484973584483

Split Value = 5.15

Left Node

Node has 36 elements of Class 0

Node has 4 elements of Class 1

Node has 1 elements of Class 2

Entropy of Left Node 0.6229837106687364

Right Node

Node has 14 elements of Class 0

Node has 46 elements of Class 1

Node has 49 elements of Class 2

Entropy of Right Node 1.424077891299181

Info gain at split 5.15 is 0.37985035212762996

Split Value = 5.25

Left Node

Node has 39 elements of Class 0

Node has 5 elements of Class 1

Node has 1 elements of Class 2

Entropy of Left Node 0.6531791627726858

Right Node

Node has 11 elements of Class 0

Node has 45 elements of Class 1  
 Node has 49 elements of Class 2  
 Entropy of Right Node 1.3779796176519241  
  
 Info gain at split 5.25 is 0.42442301953300343  
  
 Split Value = 5.35  
  
 Left Node  
 Node has 40 elements of Class 0  
 Node has 5 elements of Class 1  
 Node has 1 elements of Class 2  
 Entropy of Left Node 0.6434149067106801  
  
 Right Node  
 Node has 10 elements of Class 0  
 Node has 45 elements of Class 1  
 Node has 49 elements of Class 2  
 Entropy of Right Node 1.359348827320379  
  
 Info gain at split 5.35 is 0.4451667423877512  
  
 Split Value = 5.45  
  
 Left Node  
 Node has 45 elements of Class 0  
 Node has 6 elements of Class 1  
 Node has 1 elements of Class 2  
 Entropy of Left Node 0.6496096346956632  
  
 Right Node  
 Node has 5 elements of Class 0  
 Node has 44 elements of Class 1  
 Node has 49 elements of Class 2  
 Entropy of Right Node 1.2377158231343603  
  
 Info gain at split 5.45 is 0.5511234895788775  
  
 Split Value = 5.55  
  
 Left Node  
 Node has 47 elements of Class 0  
 Node has 11 elements of Class 1  
 Node has 1 elements of Class 2  
 Entropy of Left Node 0.8128223064150747  
  
 Right Node  
 Node has 3 elements of Class 0  
 Node has 39 elements of Class 1  
 Node has 49 elements of Class 2  
 Entropy of Right Node 1.167065448996099  
  
 Info gain at split 5.55 is 0.5572326878069267  
  
 Split Value = 5.65  
  
 Left Node  
 Node has 47 elements of Class 0

Node has 16 elements of Class 1  
Node has 2 elements of Class 2  
Entropy of Left Node 0.9905881818153937

Right Node  
Node has 3 elements of Class 0  
Node has 34 elements of Class 1  
Node has 48 elements of Class 2  
Entropy of Right Node 1.1646047697931075

Info gain at split 5.65 is 0.4957649190517246

Split Value = 5.75

Left Node  
Node has 49 elements of Class 0  
Node has 21 elements of Class 1  
Node has 3 elements of Class 2  
Entropy of Left Node 1.0923679048777892

Right Node  
Node has 1 elements of Class 0  
Node has 29 elements of Class 1  
Node has 47 elements of Class 2  
Entropy of Right Node 1.0466940096877775

Info gain at split 5.75 is 0.5160405287075729

Split Value = 5.85

Left Node  
Node has 50 elements of Class 0  
Node has 24 elements of Class 1  
Node has 6 elements of Class 2  
Entropy of Left Node 1.2251570385077257

Right Node  
Node has 26 elements of Class 1  
Node has 44 elements of Class 2  
Entropy of Right Node 0.9517626756348311

Info gain at split 5.85 is 0.48738949822078115

Split Value = 5.95

Left Node  
Node has 50 elements of Class 0  
Node has 26 elements of Class 1  
Node has 7 elements of Class 2  
Entropy of Left Node 1.2659342914094807

Right Node  
Node has 24 elements of Class 1  
Node has 43 elements of Class 2  
Entropy of Right Node 0.9411864371816836

Info gain at split 5.95 is 0.4640822508667579

Split Value = 6.05

Left Node

Node has 50 elements of Class 0

Node has 30 elements of Class 1

Node has 9 elements of Class 2

Entropy of Left Node 1.3304654268094023

Right Node

Node has 20 elements of Class 1

Node has 41 elements of Class 2

Entropy of Right Node 0.9127341558073343

Info gain at split 6.05 is 0.4243744574525947

Split Value = 6.15

Left Node

Node has 50 elements of Class 0

Node has 34 elements of Class 1

Node has 11 elements of Class 2

Entropy of Left Node 1.378063041001916

Right Node

Node has 16 elements of Class 1

Node has 39 elements of Class 2

Entropy of Right Node 0.8698926856041563

Info gain at split 6.15 is 0.39322859003175203

Split Value = 6.25

Left Node

Node has 50 elements of Class 0

Node has 36 elements of Class 1

Node has 13 elements of Class 2

Entropy of Left Node 1.4130351465796736

Right Node

Node has 14 elements of Class 1

Node has 37 elements of Class 2

Entropy of Right Node 0.8478617451660526

Info gain at split 6.25 is 0.3640863106221135

Split Value = 6.35

Left Node

Node has 50 elements of Class 0

Node has 39 elements of Class 1

Node has 19 elements of Class 2

Entropy of Left Node 1.486053069017246

Right Node

Node has 11 elements of Class 1

Node has 31 elements of Class 2

Entropy of Right Node 0.8296071030882032

Info gain at split 6.35 is 0.2827143021640419

Split Value = 6.45

Left Node

Node has 50 elements of Class 0

Node has 41 elements of Class 1

Node has 24 elements of Class 2

Entropy of Left Node 1.5246940362018935

Right Node

Node has 9 elements of Class 1

Node has 26 elements of Class 2

Entropy of Right Node 0.8224042259549891

Info gain at split 6.45 is 0.2241360869102067

Split Value = 6.55

Left Node

Node has 50 elements of Class 0

Node has 42 elements of Class 1

Node has 28 elements of Class 2

Entropy of Left Node 1.5462566034163765

Right Node

Node has 8 elements of Class 1

Node has 22 elements of Class 2

Entropy of Right Node 0.8366407419411673

Info gain at split 6.55 is 0.1806290695998214

Split Value = 6.65

Left Node

Node has 50 elements of Class 0

Node has 44 elements of Class 1

Node has 28 elements of Class 2

Entropy of Left Node 1.5453788250514908

Right Node

Node has 6 elements of Class 1

Node has 22 elements of Class 2

Entropy of Right Node 0.74959525725948

Info gain at split 6.65 is 0.1881299416575073

Split Value = 6.75

Left Node

Node has 50 elements of Class 0

Node has 47 elements of Class 1

Node has 33 elements of Class 2

Entropy of Left Node 1.562956340286807

Right Node

Node has 3 elements of Class 1

Node has 17 elements of Class 2

Entropy of Right Node 0.6098403047164004

Info gain at split 6.75 is 0.14908829851040317

Split Value = 6.85

Left Node

Node has 50 elements of Class 0

Node has 48 elements of Class 1

Node has 35 elements of Class 2

Entropy of Left Node 1.5680951037987416

Right Node

Node has 2 elements of Class 1

Node has 15 elements of Class 2

Entropy of Right Node 0.5225593745369408

Info gain at split 6.85 is 0.1353614462387518

Split Value = 6.95

Left Node

Node has 50 elements of Class 0

Node has 49 elements of Class 1

Node has 38 elements of Class 2

Entropy of Left Node 1.5744201314186457

Right Node

Node has 1 elements of Class 1

Node has 12 elements of Class 2

Entropy of Right Node 0.39124356362925566

Info gain at split 6.95 is 0.11308433851092436

Split Value = 7.05

Left Node

Node has 50 elements of Class 0

Node has 50 elements of Class 1

Node has 38 elements of Class 2

Entropy of Left Node 1.573692105413468

Right Node

Node has 12 elements of Class 2

Entropy of Right Node 0.0

Info gain at split 7.05 is 0.1371657637407655

Split Value = 7.15

Left Node

Node has 50 elements of Class 0

Node has 50 elements of Class 1

Node has 39 elements of Class 2

Entropy of Left Node 1.5756582992479669

Right Node

Node has 11 elements of Class 2



Entropy of Right Node 0.0

Info gain at split 7.15 is 0.12485247675137345

Split Value = 7.25

Left Node

Node has 50 elements of Class 0

Node has 50 elements of Class 1

Node has 42 elements of Class 2

Entropy of Left Node 1.5802897199682002

Right Node

Node has 8 elements of Class 2

Entropy of Right Node 0.0

Info gain at split 7.25 is 0.08895489915125987

Split Value = 7.35

Left Node

Node has 50 elements of Class 0

Node has 50 elements of Class 1

Node has 43 elements of Class 2

Entropy of Left Node 1.5814447393681608

Right Node

Node has 7 elements of Class 2

Entropy of Right Node 0.0

Info gain at split 7.35 is 0.0773185158568428

Split Value = 7.5

Left Node

Node has 50 elements of Class 0

Node has 50 elements of Class 1

Node has 44 elements of Class 2

Entropy of Left Node 1.5824207639595793

Right Node

Node has 6 elements of Class 2

Entropy of Right Node 0.0

Info gain at split 7.5 is 0.06583856731995996

Split Value = 7.65

Left Node

Node has 50 elements of Class 0

Node has 50 elements of Class 1

Node has 45 elements of Class 2

Entropy of Left Node 1.5832262740679837

Right Node

Node has 5 elements of Class 2

Entropy of Right Node 0.0

Info gain at split 7.65 is 0.054510435788771794

Split Value = 7.800000000000001

Left Node

Node has 50 elements of Class 0

Node has 50 elements of Class 1

Node has 49 elements of Class 2

Entropy of Left Node 1.5848973705351974

Right Node

Node has 1 elements of Class 2

Entropy of Right Node 0.0

Info gain at split 7.800000000000001 is 0.01063111265619332

```
info_gains.index(max(info_gains))
```

```
12
```

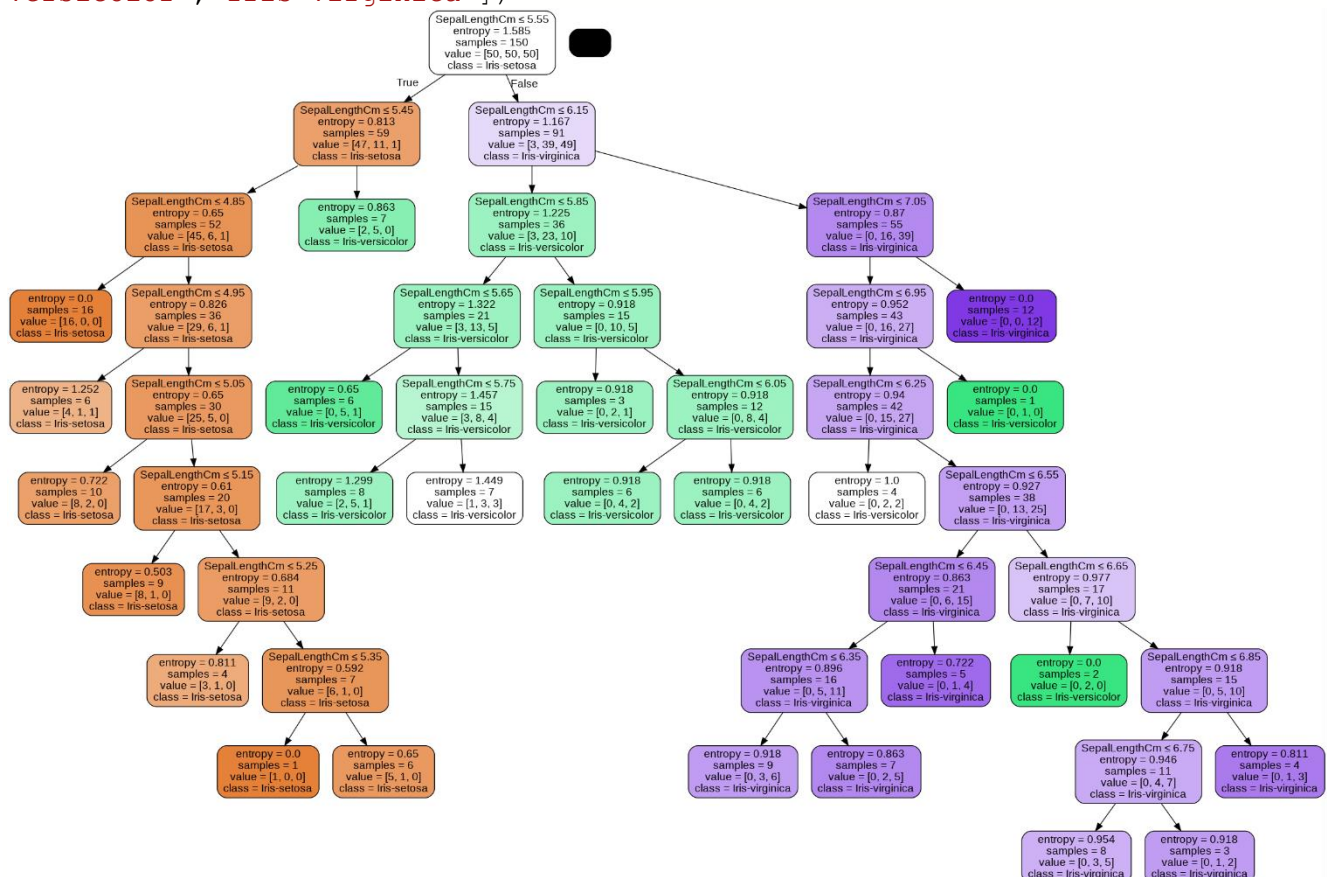
```
split_values[info_gains.index(max(info_gains))]
```

```
5.55
```

```
clf = tree.DecisionTreeClassifier(criterion='entropy')
```

```
f = clf.fit(X_train,Y_train)
```

```
plot_decision_tree(clf, X_train.columns, ["Iris-setosa", "Iris-  
versicolor", "Iris-virginica"])
```



```

X_train = df[['SepalWidthCm']]
Y_train = df.Species
entropy_before_split = entropy(df.Species)
print ("Initial Entropy = {}".format(entropy_before_split))
a=df['SepalWidthCm'].to_numpy()
a=np.unique(a)
a=np.sort(a)
l=[]
for i in range(len(a)-1):
    k=(a[i]+a[i+1])/2
    l.append(k)
split_values=np.array(l)
info_gains = []
for sp_value in split_values:
    left_data = df.loc[df.SepalWidthCm < sp_value]
    right_data = df.loc[df.SepalWidthCm > sp_value]
    info_gains.append(
        calc_info_gain_numeric(
            df, left_data.Species, right_data.Species, entropy_before_split, s
p_value
        )
    )
Node has 50 elements of Class 0
Node has 50 elements of Class 1
Node has 50 elements of Class 2
Initial Entropy = 1.584962500721156
Split Value = 2.1

Left Node
Node has 1 elements of Class 1
Entropy of Left Node 0.0

Right Node
Node has 50 elements of Class 0
Node has 49 elements of Class 1
Node has 50 elements of Class 2
Entropy of Right Node 1.5848973705351974

Info gain at split 2.1 is 0.01063111265619332

Split Value = 2.25

Left Node
Node has 3 elements of Class 1
Node has 1 elements of Class 2
Entropy of Left Node 0.8112781244591328

Right Node
Node has 50 elements of Class 0
Node has 47 elements of Class 1
Node has 49 elements of Class 2
Entropy of Right Node 1.5844871076524119

```

Info gain at split 2.25 is 0.02109429928723161

Split Value = 2.3499999999999996

Left Node

Node has 1 elements of Class 0

Node has 6 elements of Class 1

Node has 1 elements of Class 2

Entropy of Left Node 1.061278124459133

Right Node

Node has 49 elements of Class 0

Node has 44 elements of Class 1

Node has 49 elements of Class 2

Entropy of Right Node 1.583151655115024

Info gain at split 2.3499999999999996 is 0.02964410057444633

Split Value = 2.45

Left Node

Node has 1 elements of Class 0

Node has 9 elements of Class 1

Node has 1 elements of Class 2

Entropy of Left Node 0.8658566174572235

Right Node

Node has 49 elements of Class 0

Node has 41 elements of Class 1

Node has 49 elements of Class 2

Entropy of Right Node 1.5800835409779546

Info gain at split 2.45 is 0.05725560080138847

Split Value = 2.55

Left Node

Node has 1 elements of Class 0

Node has 13 elements of Class 1

Node has 5 elements of Class 2

Entropy of Left Node 1.1050139971135322

Right Node

Node has 49 elements of Class 0

Node has 37 elements of Class 1

Node has 45 elements of Class 2

Entropy of Right Node 1.5753776105088113

Info gain at split 2.55 is 0.06916428124241358

Split Value = 2.6500000000000004

Left Node

Node has 1 elements of Class 0

Node has 16 elements of Class 1

Node has 7 elements of Class 2

Entropy of Left Node 1.0994839817876882

Right Node  
 Node has 49 elements of Class 0  
 Node has 34 elements of Class 1  
 Node has 43 elements of Class 2  
 Entropy of Right Node 1.5691537059098613  
  
 Info gain at split 2.6500000000000004 is 0.09095595067084261  
  
 Split Value = 2.75  
  
 Left Node  
 Node has 1 elements of Class 0  
 Node has 21 elements of Class 1  
 Node has 11 elements of Class 2  
 Entropy of Left Node 1.0961391592565068  
  
 Right Node  
 Node has 49 elements of Class 0  
 Node has 29 elements of Class 1  
 Node has 39 elements of Class 2  
 Entropy of Right Node 1.5529893541477424  
  
 Info gain at split 2.75 is 0.13248018944948536  
  
 Split Value = 2.8499999999999996  
  
 Left Node  
 Node has 1 elements of Class 0  
 Node has 27 elements of Class 1  
 Node has 19 elements of Class 2  
 Entropy of Left Node 1.1058104407448335  
  
 Right Node  
 Node has 49 elements of Class 0  
 Node has 23 elements of Class 1  
 Node has 31 elements of Class 2  
 Entropy of Right Node 1.514240401232251  
  
 Info gain at split 2.8499999999999996 is 0.19869682044162928  
  
 Split Value = 2.95  
  
 Left Node  
 Node has 2 elements of Class 0  
 Node has 34 elements of Class 1  
 Node has 21 elements of Class 2  
 Entropy of Left Node 1.1449531285355223  
  
 Right Node  
 Node has 48 elements of Class 0  
 Node has 16 elements of Class 1  
 Node has 29 elements of Class 2  
 Entropy of Right Node 1.453571188599267  
  
 Info gain at split 2.95 is 0.2486661749461121  
  
 Split Value = 3.05

Left Node  
Node has 8 elements of Class 0  
Node has 42 elements of Class 1  
Node has 33 elements of Class 2  
Entropy of Left Node 1.3516377723645525

Right Node  
Node has 42 elements of Class 0  
Node has 8 elements of Class 1  
Node has 17 elements of Class 2  
Entropy of Right Node 1.2905041149657768

Info gain at split 3.05 is 0.2606310953280566

Split Value = 3.1500000000000004

Left Node  
Node has 13 elements of Class 0  
Node has 45 elements of Class 1  
Node has 37 elements of Class 2  
Entropy of Left Node 1.4331305536052563

Right Node  
Node has 37 elements of Class 0  
Node has 5 elements of Class 1  
Node has 13 elements of Class 2  
Entropy of Right Node 1.1910845982784386

Info gain at split 3.1500000000000004 is 0.24058213073573298

Split Value = 3.25

Left Node  
Node has 18 elements of Class 0  
Node has 48 elements of Class 1  
Node has 42 elements of Class 2  
Entropy of Left Node 1.4806821149663847

Right Node  
Node has 32 elements of Class 0  
Node has 2 elements of Class 1  
Node has 8 elements of Class 2  
Entropy of Right Node 0.9637459942073318

Info gain at split 3.25 is 0.24902249956730627

Split Value = 3.3499999999999996

Left Node  
Node has 20 elements of Class 0  
Node has 49 elements of Class 1  
Node has 45 elements of Class 2  
Entropy of Left Node 1.4934800703557258

Right Node  
Node has 30 elements of Class 0  
Node has 1 elements of Class 1

Node has 5 elements of Class 2  
Entropy of Right Node 0.7583594919230799

Info gain at split 3.3499999999999996 is 0.26791136918926517

Split Value = 3.45

Left Node  
Node has 29 elements of Class 0  
Node has 50 elements of Class 1  
Node has 47 elements of Class 2  
Entropy of Left Node 1.5475979046405137

Right Node  
Node has 21 elements of Class 0  
Node has 3 elements of Class 2  
Entropy of Right Node 0.5435644431995964

Info gain at split 3.45 is 0.19800994991118914

Split Value = 3.55

Left Node  
Node has 35 elements of Class 0  
Node has 50 elements of Class 1  
Node has 47 elements of Class 2  
Entropy of Left Node 1.5687623685201277

Right Node  
Node has 15 elements of Class 0  
Node has 3 elements of Class 2  
Entropy of Right Node 0.6500224216483541

Info gain at split 3.55 is 0.12644892582564116

Split Value = 3.6500000000000004

Left Node  
Node has 37 elements of Class 0  
Node has 50 elements of Class 1  
Node has 48 elements of Class 2  
Entropy of Left Node 1.5729579374091802

Right Node  
Node has 13 elements of Class 0  
Node has 2 elements of Class 2  
Entropy of Right Node 0.5665095065529053

Info gain at split 3.6500000000000004 is 0.11264940639760335

Split Value = 3.75

Left Node  
Node has 40 elements of Class 0  
Node has 50 elements of Class 1  
Node has 48 elements of Class 2  
Entropy of Left Node 1.5784655196850803

Right Node  
Node has 10 elements of Class 0  
Node has 2 elements of Class 2  
Entropy of Right Node 0.6500224216483541

Info gain at split 3.75 is 0.0807724288790137

Split Value = 3.8499999999999996

Left Node  
Node has 44 elements of Class 0  
Node has 50 elements of Class 1  
Node has 50 elements of Class 2  
Entropy of Left Node 1.5824207639595793

Right Node  
Node has 6 elements of Class 0  
Entropy of Right Node 0.0

Info gain at split 3.8499999999999996 is 0.06583856731995996

Split Value = 3.95

Left Node  
Node has 46 elements of Class 0  
Node has 50 elements of Class 1  
Node has 50 elements of Class 2  
Entropy of Left Node 1.5838692920574484

Right Node  
Node has 4 elements of Class 0  
Entropy of Right Node 0.0

Info gain at split 3.95 is 0.04332972311857297

Split Value = 4.05

Left Node  
Node has 47 elements of Class 0  
Node has 50 elements of Class 1  
Node has 50 elements of Class 2  
Entropy of Left Node 1.5843574128205722

Right Node  
Node has 3 elements of Class 0  
Entropy of Right Node 0.0

Info gain at split 4.05 is 0.03229223615699528

Split Value = 4.15

Left Node  
Node has 48 elements of Class 0  
Node has 50 elements of Class 1  
Node has 50 elements of Class 2  
Entropy of Left Node 1.5846978317634905

Right Node



Node has 2 elements of Class 0  
Entropy of Right Node 0.0

Info gain at split 4.15 is 0.021393973381178766

Split Value = 4.300000000000001

Left Node

Node has 49 elements of Class 0

Node has 50 elements of Class 1

Node has 50 elements of Class 2

Entropy of Left Node 1.5848973705351974

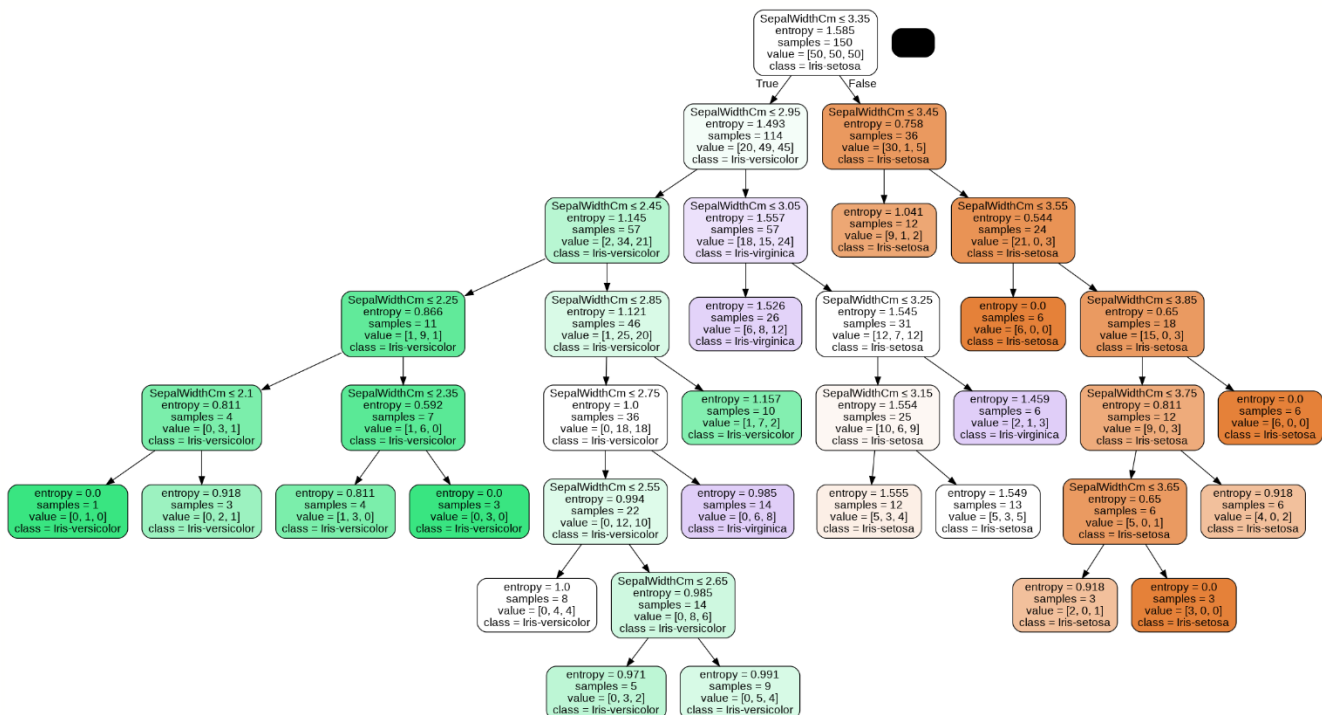
Right Node

Node has 1 elements of Class 0

Entropy of Right Node 0.0

Info gain at split 4.300000000000001 is 0.01063111265619332

```
print(info_gains.index(max(info_gains)))  
split_values[info_gains.index(max(info_gains))]  
clf = tree.DecisionTreeClassifier(criterion='entropy')  
clf = clf.fit(X_train,Y_train)  
plot_decision_tree(clf, X_train.columns, ["Iris-setosa", "Iris-versicol  
or", "Iris-virginica"])  
12
```



```
X_train = df[['PetalLengthCm']]
```

```
Y_train = df.Species
```

```
entropy_before_split = entropy(df.Species)
```

```
print ("Initial Entropy = {}".format(entropy_before_split))
```

```
a=df['PetalLengthCm'].to_numpy()
```

```

a=np.unique(a)
a=np.sort(a)
l=[]
for i in range(len(a)-1):
    k=(a[i]+a[i+1])/2
    l.append(k)
split_values=np.array(l)
info_gains = []
for sp_value in split_values:
    left_data = df.loc[df.PetalLengthCm < sp_value]
    right_data = df.loc[df.PetalLengthCm > sp_value]
    info_gains.append(
        calc_info_gain_numeric(
            df, left_data.Species, right_data.Species, entropy_before_split, s
p_value
        )
    )

```

```

Node has 50 elements of Class 0
Node has 50 elements of Class 1
Node has 50 elements of Class 2
Initial Entropy = 1.584962500721156
Split Value = 1.05

```

```

Left Node
Node has 1 elements of Class 0
Entropy of Left Node 0.0

```

```

Right Node
Node has 49 elements of Class 0
Node has 50 elements of Class 1
Node has 50 elements of Class 2
Entropy of Right Node 1.5848973705351974

```

```

Info gain at split 1.05 is 0.01063111265619332

```

```

Split Value = 1.15

```

```

Left Node
Node has 2 elements of Class 0
Entropy of Left Node 0.0

```

```

Right Node
Node has 48 elements of Class 0
Node has 50 elements of Class 1
Node has 50 elements of Class 2
Entropy of Right Node 1.5846978317634905

```

```

Info gain at split 1.15 is 0.021393973381178766

```

```

Split Value = 1.25

```

```

Left Node
Node has 4 elements of Class 0

```

Entropy of Left Node 0.0

Right Node

Node has 46 elements of Class 0

Node has 50 elements of Class 1

Node has 50 elements of Class 2

Entropy of Right Node 1.5838692920574484

Info gain at split 1.25 is 0.04332972311857297

Split Value = 1.35

Left Node

Node has 11 elements of Class 0

Entropy of Left Node 0.0

Right Node

Node has 39 elements of Class 0

Node has 50 elements of Class 1

Node has 50 elements of Class 2

Entropy of Right Node 1.5756582992479666

Info gain at split 1.35 is 0.12485247675137368

Split Value = 1.45

Left Node

Node has 23 elements of Class 0

Entropy of Left Node 0.0

Right Node

Node has 27 elements of Class 0

Node has 50 elements of Class 1

Node has 50 elements of Class 2

Entropy of Right Node 1.5338218400329051

Info gain at split 1.45 is 0.286326676159963

Split Value = 1.55

Left Node

Node has 37 elements of Class 0

Entropy of Left Node 0.0

Right Node

Node has 13 elements of Class 0

Node has 50 elements of Class 1

Node has 50 elements of Class 2

Entropy of Right Node 1.3999016587576067

Info gain at split 1.55 is 0.5303699177904257

Split Value = 1.65

Left Node

Node has 44 elements of Class 0

Entropy of Left Node 0.0

Right Node  
 Node has 6 elements of Class 0  
 Node has 50 elements of Class 1  
 Node has 50 elements of Class 2  
 Entropy of Right Node 1.2572091905839593  
  
 Info gain at split 1.65 is 0.6965346727084915  
  
 Split Value = 1.7999999999999998  
  
 Left Node  
 Node has 48 elements of Class 0  
 Entropy of Left Node 0.0  
  
 Right Node  
 Node has 2 elements of Class 0  
 Node has 50 elements of Class 1  
 Node has 50 elements of Class 2  
 Entropy of Right Node 1.119625155917844  
  
 Info gain at split 1.7999999999999998 is 0.8236173946970221  
  
 Split Value = 2.45  
  
 Left Node  
 Node has 50 elements of Class 0  
 Entropy of Left Node 0.0  
  
 Right Node  
 Node has 50 elements of Class 1  
 Node has 50 elements of Class 2  
 Entropy of Right Node 1.0  
  
 Info gain at split 2.45 is 0.9182958340544894  
  
 Split Value = 3.15  
  
 Left Node  
 Node has 50 elements of Class 0  
 Node has 1 elements of Class 1  
 Entropy of Left Node 0.13923299905509887  
  
 Right Node  
 Node has 49 elements of Class 1  
 Node has 50 elements of Class 2  
 Entropy of Right Node 0.999926399368686  
  
 Info gain at split 3.15 is 0.8776718574590896  
  
 Split Value = 3.4  
  
 Left Node  
 Node has 50 elements of Class 0  
 Node has 3 elements of Class 1  
 Entropy of Left Node 0.3138129641688651  
  
 Right Node  
 Node has 47 elements of Class 1

Node has 50 elements of Class 2  
Entropy of Right Node 0.999309898706868

Info gain at split 3.4 is 0.8278615188843824

Split Value = 3.55

Left Node  
Node has 50 elements of Class 0  
Node has 5 elements of Class 1  
Entropy of Left Node 0.4394969869215134

Right Node  
Node has 45 elements of Class 1  
Node has 50 elements of Class 2  
Entropy of Right Node 0.9980008838722996

Info gain at split 3.55 is 0.7917463790641448

Split Value = 3.6500000000000004

Left Node  
Node has 50 elements of Class 0  
Node has 6 elements of Class 1  
Entropy of Left Node 0.49123734182433315

Right Node  
Node has 44 elements of Class 1  
Node has 50 elements of Class 2  
Entropy of Right Node 0.9970590569034106

Info gain at split 3.6500000000000004 is 0.7767435507806011

Split Value = 3.75

Left Node  
Node has 50 elements of Class 0  
Node has 7 elements of Class 1  
Entropy of Left Node 0.5373760853377336

Right Node  
Node has 43 elements of Class 1  
Node has 50 elements of Class 2  
Entropy of Right Node 0.9959094138937685

Info gain at split 3.75 is 0.7632957516786808

Split Value = 3.8499999999999996

Left Node  
Node has 50 elements of Class 0  
Node has 8 elements of Class 1  
Entropy of Left Node 0.5787946246321198

Right Node  
Node has 42 elements of Class 1  
Node has 50 elements of Class 2  
Entropy of Right Node 0.994538681650011

Info gain at split 3.8499999999999996 is 0.7511781877847297

Split Value = 3.95

Left Node

Node has 50 elements of Class 0

Node has 11 elements of Class 1

Entropy of Left Node 0.6807937753703206

Right Node

Node has 39 elements of Class 1

Node has 50 elements of Class 2

Entropy of Right Node 0.9889525767600615

Info gain at split 3.95 is 0.7213278365262558

Split Value = 4.05

Left Node

Node has 50 elements of Class 0

Node has 16 elements of Class 1

Entropy of Left Node 0.7990485210442682

Right Node

Node has 34 elements of Class 1

Node has 50 elements of Class 2

Entropy of Right Node 0.9736680645496202

Info gain at split 4.05 is 0.6881270353138906

Split Value = 4.15

Left Node

Node has 50 elements of Class 0

Node has 19 elements of Class 1

Entropy of Left Node 0.8490544242540479

Right Node

Node has 31 elements of Class 1

Node has 50 elements of Class 2

Entropy of Right Node 0.9599377175669781

Info gain at split 4.15 is 0.6760310980781259

Split Value = 4.25

Left Node

Node has 50 elements of Class 0

Node has 23 elements of Class 1

Entropy of Left Node 0.8989377852081333

Right Node

Node has 27 elements of Class 1

Node has 50 elements of Class 2

Entropy of Right Node 0.9346466439786691

Info gain at split 4.25 is 0.6676941680108144

Split Value = 4.35

Left Node

Node has 50 elements of Class 0

Node has 25 elements of Class 1

Entropy of Left Node 0.9182958340544896

Right Node

Node has 25 elements of Class 1

Node has 50 elements of Class 2

Entropy of Right Node 0.9182958340544896

Info gain at split 4.35 is 0.6666666666666665

Split Value = 4.45

Left Node

Node has 50 elements of Class 0

Node has 29 elements of Class 1

Entropy of Left Node 0.9484103893488013

Right Node

Node has 21 elements of Class 1

Node has 50 elements of Class 2

Entropy of Right Node 0.8760643678555242

Info gain at split 4.45 is 0.6707958948791727

Split Value = 4.55

Left Node

Node has 50 elements of Class 0

Node has 36 elements of Class 1

Node has 1 elements of Class 2

Entropy of Left Node 1.0600686735422975

Right Node

Node has 14 elements of Class 1

Node has 49 elements of Class 2

Entropy of Right Node 0.7642045065086203

Info gain at split 4.55 is 0.6491567773330031

Split Value = 4.65

Left Node

Node has 50 elements of Class 0

Node has 39 elements of Class 1

Node has 1 elements of Class 2

Entropy of Left Node 1.0660364738367423

Right Node

Node has 11 elements of Class 1

Node has 49 elements of Class 2

Entropy of Right Node 0.6873150928309273

Info gain at split 4.65 is 0.6704145792867398

Split Value = 4.75

Left Node

Node has 50 elements of Class 0

Node has 44 elements of Class 1

Node has 1 elements of Class 2

Entropy of Left Node 1.0708261271859234

Right Node

Node has 6 elements of Class 1

Node has 49 elements of Class 2

Entropy of Right Node 0.4971677614160753

Info gain at split 4.75 is 0.7244777743175104

Split Value = 4.85

Left Node

Node has 50 elements of Class 0

Node has 46 elements of Class 1

Node has 3 elements of Class 2

Entropy of Left Node 1.1643914991753441

Right Node

Node has 4 elements of Class 1

Node has 47 elements of Class 2

Entropy of Right Node 0.39662777277837885

Info gain at split 4.85 is 0.6816106685207801

Split Value = 4.95

Left Node

Node has 50 elements of Class 0

Node has 48 elements of Class 1

Node has 6 elements of Class 2

Entropy of Left Node 1.2602398669134125

Right Node

Node has 2 elements of Class 1

Node has 44 elements of Class 2

Entropy of Right Node 0.2580186686648155

Info gain at split 4.95 is 0.6320704679373134

Split Value = 5.05

Left Node

Node has 50 elements of Class 0

Node has 49 elements of Class 1

Node has 9 elements of Class 2

Entropy of Left Node 1.330416049724929

Right Node

Node has 1 elements of Class 1

Node has 41 elements of Class 2

Entropy of Right Node 0.1623261801753929



Info gain at split 5.05 is 0.5816116144700971

Split Value = 5.15

Left Node

Node has 50 elements of Class 0

Node has 50 elements of Class 1

Node has 16 elements of Class 2

Entropy of Left Node 1.4408635901493612

Right Node

Node has 34 elements of Class 2

Entropy of Right Node 0.0

Info gain at split 5.15 is 0.4706946576723168

Split Value = 5.25

Left Node

Node has 50 elements of Class 0

Node has 50 elements of Class 1

Node has 18 elements of Class 2

Entropy of Left Node 1.4636238205191796

Right Node

Node has 32 elements of Class 2

Entropy of Right Node 0.0

Info gain at split 5.25 is 0.43357842857940154

Split Value = 5.35

Left Node

Node has 50 elements of Class 0

Node has 50 elements of Class 1

Node has 20 elements of Class 2

Entropy of Left Node 1.4833557549816876

Right Node

Node has 30 elements of Class 2

Entropy of Right Node 0.0

Info gain at split 5.35 is 0.39827789673580605

Split Value = 5.45

Left Node

Node has 50 elements of Class 0

Node has 50 elements of Class 1

Node has 22 elements of Class 2

Entropy of Left Node 1.5004659065178614

Right Node

Node has 28 elements of Class 2

Entropy of Right Node 0.0

Info gain at split 5.45 is 0.3645835634199621

Split Value = 5.55

Left Node

Node has 50 elements of Class 0

Node has 50 elements of Class 1

Node has 25 elements of Class 2

Entropy of Left Node 1.5219280948873621

Right Node

Node has 25 elements of Class 2

Entropy of Right Node 0.0

Info gain at split 5.55 is 0.31668908831502085

Split Value = 5.65

Left Node

Node has 50 elements of Class 0

Node has 50 elements of Class 1

Node has 31 elements of Class 2

Entropy of Left Node 1.5527687679537436

Right Node

Node has 19 elements of Class 2

Entropy of Right Node 0.0

Info gain at split 5.65 is 0.22887777670821996

Split Value = 5.75

Left Node

Node has 50 elements of Class 0

Node has 50 elements of Class 1

Node has 34 elements of Class 2

Entropy of Left Node 1.5634074323981904

Right Node

Node has 16 elements of Class 2

Entropy of Right Node 0.0

Info gain at split 5.75 is 0.18831852777877267

Split Value = 5.85

Left Node

Node has 50 elements of Class 0

Node has 50 elements of Class 1

Node has 37 elements of Class 2

Entropy of Left Node 1.5714963639405002

Right Node

Node has 13 elements of Class 2

Entropy of Right Node 0.0

Info gain at split 5.85 is 0.149662488322166

Split Value = 5.95

Left Node  
Node has 50 elements of Class 0  
Node has 50 elements of Class 1  
Node has 39 elements of Class 2  
Entropy of Left Node 1.5756582992479669

Right Node  
Node has 11 elements of Class 2  
Entropy of Right Node 0.0

Info gain at split 5.95 is 0.12485247675137345

Split Value = 6.05

Left Node  
Node has 50 elements of Class 0  
Node has 50 elements of Class 1  
Node has 41 elements of Class 2  
Entropy of Left Node 1.5789467341944396

Right Node  
Node has 9 elements of Class 2  
Entropy of Right Node 0.0

Info gain at split 6.05 is 0.10075257057838294

Split Value = 6.199999999999999

Left Node  
Node has 50 elements of Class 0  
Node has 50 elements of Class 1  
Node has 44 elements of Class 2  
Entropy of Left Node 1.5824207639595793

Right Node  
Node has 6 elements of Class 2  
Entropy of Right Node 0.0

Info gain at split 6.199999999999999 is 0.06583856731995996

Split Value = 6.35

Left Node  
Node has 50 elements of Class 0  
Node has 50 elements of Class 1  
Node has 45 elements of Class 2  
Entropy of Left Node 1.5832262740679837

Right Node  
Node has 5 elements of Class 2  
Entropy of Right Node 0.0

Info gain at split 6.35 is 0.054510435788771794

Split Value = 6.5

Left Node

Node has 50 elements of Class 0  
Node has 50 elements of Class 1  
Node has 46 elements of Class 2  
Entropy of Left Node 1.5838692920574484

Right Node  
Node has 4 elements of Class 2  
Entropy of Right Node 0.0

Info gain at split 6.5 is 0.04332972311857297

Split Value = 6.65

Left Node  
Node has 50 elements of Class 0  
Node has 50 elements of Class 1  
Node has 47 elements of Class 2  
Entropy of Left Node 1.5843574128205722

Right Node  
Node has 3 elements of Class 2  
Entropy of Right Node 0.0

Info gain at split 6.65 is 0.03229223615699528

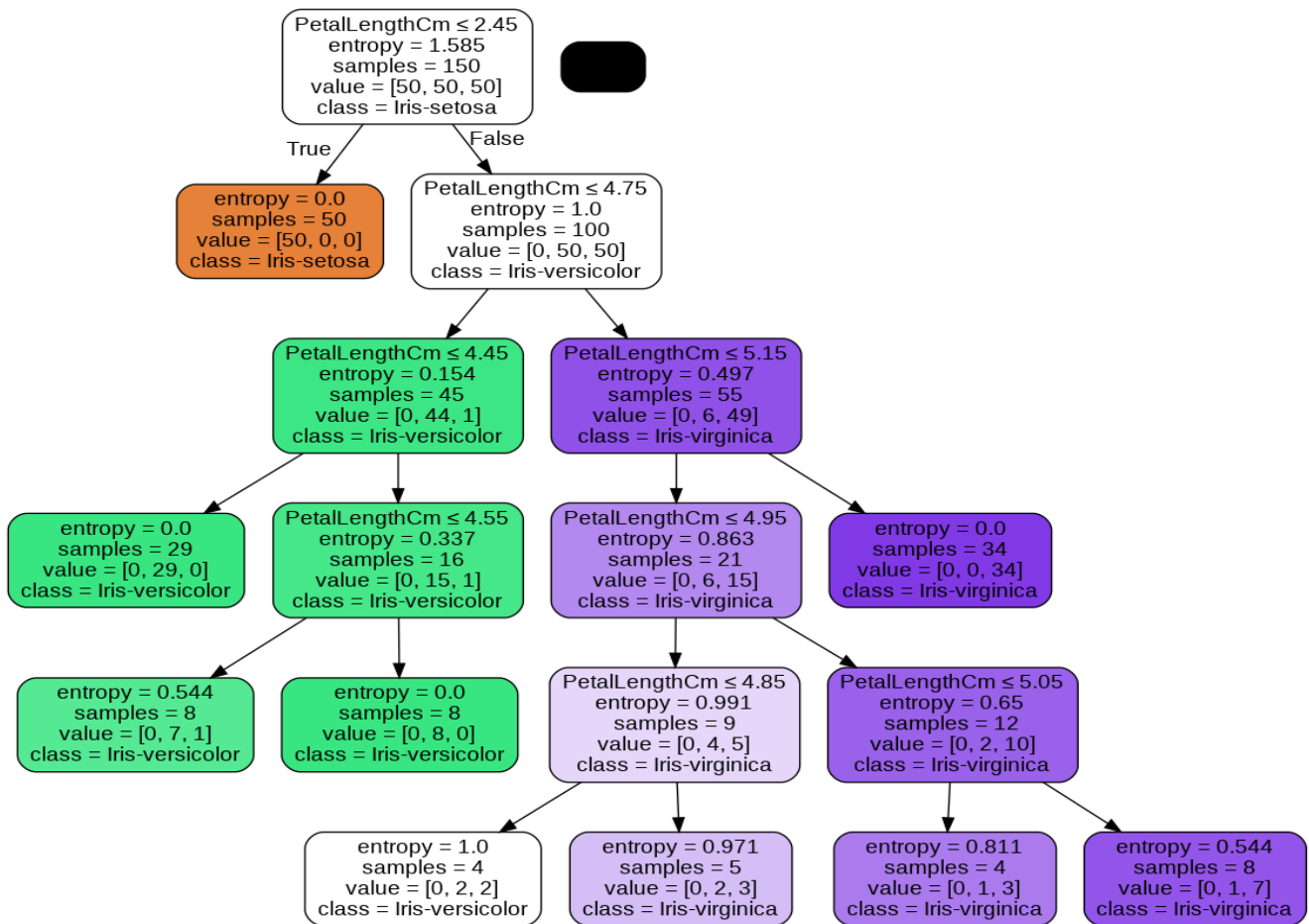
Split Value = 6.8000000000000001

Left Node  
Node has 50 elements of Class 0  
Node has 50 elements of Class 1  
Node has 49 elements of Class 2  
Entropy of Left Node 1.5848973705351974

Right Node  
Node has 1 elements of Class 2  
Entropy of Right Node 0.0

Info gain at split 6.8000000000000001 is 0.01063111265619332

```
info_gains.index(max(info_gains))  
split_values[info_gains.index(max(info_gains))]  
clf = tree.DecisionTreeClassifier(criterion='entropy')  
clf = clf.fit(X_train,Y_train)  
plot_decision_tree(clf, X_train.columns, ["Iris-setosa", "Iris-  
versicolor", "Iris-virginica"])
```



```

X_train = df[['PetalWidthCm']]
Y_train = df.Species
entropy_before_split = entropy(df.Species)
print ("Initial Entropy = {}".format(entropy_before_split))
a=df['PetalWidthCm'].to_numpy()
a=np.unique(a)
a=np.sort(a)
l=[]
for i in range(len(a)-1):
    k=(a[i]+a[i+1])/2
    l.append(k)
split_values=np.array(l)
info_gains = []
for sp_value in split_values:
    left_data = df.loc[df.PetalWidthCm < sp_value]
    right_data = df.loc[df.PetalWidthCm > sp_value]
    info_gains.append(
        calc_info_gain_numeric(
            df, left_data.Species, right_data.Species, entropy_before_split, s
p_value
        )
    )

```

Node has 50 elements of Class 0  
Node has 50 elements of Class 1  
Node has 50 elements of Class 2  
Initial Entropy = 1.584962500721156  
Split Value = 0.15000000000000002

Left Node  
Node has 6 elements of Class 0  
Entropy of Left Node 0.0

Right Node  
Node has 44 elements of Class 0  
Node has 50 elements of Class 1  
Node has 50 elements of Class 2  
Entropy of Right Node 1.5824207639595793

Info gain at split 0.15000000000000002 is 0.06583856731995996

Split Value = 0.25

Left Node  
Node has 34 elements of Class 0  
Entropy of Left Node 0.0

Right Node  
Node has 16 elements of Class 0  
Node has 50 elements of Class 1  
Node has 50 elements of Class 2  
Entropy of Right Node 1.4408635901493612

Info gain at split 0.25 is 0.4706946576723168

Split Value = 0.35

Left Node  
Node has 41 elements of Class 0  
Entropy of Left Node 0.0

Right Node  
Node has 9 elements of Class 0  
Node has 50 elements of Class 1  
Node has 50 elements of Class 2  
Entropy of Right Node 1.3285976826626762

Info gain at split 0.35 is 0.6195148513196114

Split Value = 0.45

Left Node  
Node has 48 elements of Class 0  
Entropy of Left Node 0.0

Right Node  
Node has 2 elements of Class 0  
Node has 50 elements of Class 1  
Node has 50 elements of Class 2  
Entropy of Right Node 1.119625155917844

Info gain at split 0.45 is 0.8236173946970221

Split Value = 0.55

Left Node

Node has 49 elements of Class 0

Entropy of Left Node 0.0

Right Node

Node has 1 elements of Class 0

Node has 50 elements of Class 1

Node has 50 elements of Class 2

Entropy of Right Node 1.0702350572322654

Info gain at split 0.55 is 0.864337562184764

Split Value = 0.8

Left Node

Node has 50 elements of Class 0

Entropy of Left Node 0.0

Right Node

Node has 50 elements of Class 1

Node has 50 elements of Class 2

Entropy of Right Node 1.0

Info gain at split 0.8 is 0.9182958340544894

Split Value = 1.05

Left Node

Node has 50 elements of Class 0

Node has 7 elements of Class 1

Entropy of Left Node 0.5373760853377336

Right Node

Node has 43 elements of Class 1

Node has 50 elements of Class 2

Entropy of Right Node 0.9959094138937685

Info gain at split 1.05 is 0.7632957516786808

Split Value = 1.15

Left Node

Node has 50 elements of Class 0

Node has 10 elements of Class 1

Entropy of Left Node 0.6500224216483541

Right Node

Node has 40 elements of Class 1

Node has 50 elements of Class 2

Entropy of Right Node 0.9910760598382222

Info gain at split 1.15 is 0.7303078961588811

Split Value = 1.25

Left Node  
 Node has 50 elements of Class 0  
 Node has 15 elements of Class 1  
 Entropy of Left Node 0.7793498372920852  
  
 Right Node  
 Node has 35 elements of Class 1  
 Node has 50 elements of Class 2  
 Entropy of Right Node 0.9774178175281716  
  
 Info gain at split 1.25 is 0.6933741412952886  
  
 Split Value = 1.35  
  
 Left Node  
 Node has 50 elements of Class 0  
 Node has 28 elements of Class 1  
 Entropy of Left Node 0.9418285354475155  
  
 Right Node  
 Node has 22 elements of Class 1  
 Node has 50 elements of Class 2  
 Entropy of Right Node 0.8879763195151349  
  
 Info gain at split 1.35 is 0.6689830289211832  
  
 Split Value = 1.45  
  
 Left Node  
 Node has 50 elements of Class 0  
 Node has 35 elements of Class 1  
 Node has 1 elements of Class 2  
 Entropy of Left Node 1.0574541142159344  
  
 Right Node  
 Node has 15 elements of Class 1  
 Node has 49 elements of Class 2  
 Entropy of Right Node 0.7855602922535472  
  
 Info gain at split 1.45 is 0.6435164172091734  
  
 Split Value = 1.55  
  
 Left Node  
 Node has 50 elements of Class 0  
 Node has 45 elements of Class 1  
 Node has 3 elements of Class 2  
 Entropy of Left Node 1.1649028407913813  
  
 Right Node  
 Node has 5 elements of Class 1  
 Node has 47 elements of Class 2  
 Entropy of Right Node 0.4566836315394428  
  
 Info gain at split 1.55 is 0.6655756524704468  
  
 Split Value = 1.65



Left Node  
 Node has 50 elements of Class 0  
 Node has 48 elements of Class 1  
 Node has 4 elements of Class 2  
 Entropy of Left Node 1.1991801505660866

Right Node  
 Node has 2 elements of Class 1  
 Node has 46 elements of Class 2  
 Entropy of Right Node 0.24988229283318544

Info gain at split 1.65 is 0.6895576646295978

Split Value = 1.75

Left Node  
 Node has 50 elements of Class 0  
 Node has 49 elements of Class 1  
 Node has 5 elements of Class 2  
 Entropy of Left Node 1.2300240227101475

Right Node  
 Node has 1 elements of Class 1  
 Node has 45 elements of Class 2  
 Entropy of Right Node 0.15109697051711368

Info gain at split 1.75 is 0.6858094406835389

Split Value = 1.85

Left Node  
 Node has 50 elements of Class 0  
 Node has 50 elements of Class 1  
 Node has 16 elements of Class 2  
 Entropy of Left Node 1.4408635901493612

Right Node  
 Node has 34 elements of Class 2  
 Entropy of Right Node 0.0

Info gain at split 1.85 is 0.4706946576723168

Split Value = 1.95

Left Node  
 Node has 50 elements of Class 0  
 Node has 50 elements of Class 1  
 Node has 21 elements of Class 2  
 Entropy of Left Node 1.492216254995037

Right Node  
 Node has 29 elements of Class 2  
 Entropy of Right Node 0.0

Info gain at split 1.95 is 0.38124138835849286

Split Value = 2.05

```

Left Node
Node has 50 elements of Class 0
Node has 50 elements of Class 1
Node has 27 elements of Class 2
Entropy of Left Node 1.5338218400329051

Right Node
Node has 23 elements of Class 2
Entropy of Right Node 0.0

Info gain at split 2.05 is 0.286326676159963

Split Value = 2.1500000000000004

Left Node
Node has 50 elements of Class 0
Node has 50 elements of Class 1
Node has 33 elements of Class 2
Entropy of Left Node 1.5601649549274947

Right Node
Node has 17 elements of Class 2
Entropy of Right Node 0.0

Info gain at split 2.1500000000000004 is 0.20161624068544404

Split Value = 2.25

Left Node
Node has 50 elements of Class 0
Node has 50 elements of Class 1
Node has 36 elements of Class 2
Entropy of Left Node 1.5690590248577239

Right Node
Node has 14 elements of Class 2
Entropy of Right Node 0.0

Info gain at split 2.25 is 0.16234898485015314

Split Value = 2.3499999999999996

Left Node
Node has 50 elements of Class 0
Node has 50 elements of Class 1
Node has 44 elements of Class 2
Entropy of Left Node 1.5824207639595793

Right Node
Node has 6 elements of Class 2
Entropy of Right Node 0.0

Info gain at split 2.3499999999999996 is 0.06583856731995996
Split Value = 2.45

Left Node
Node has 50 elements of Class 0

```

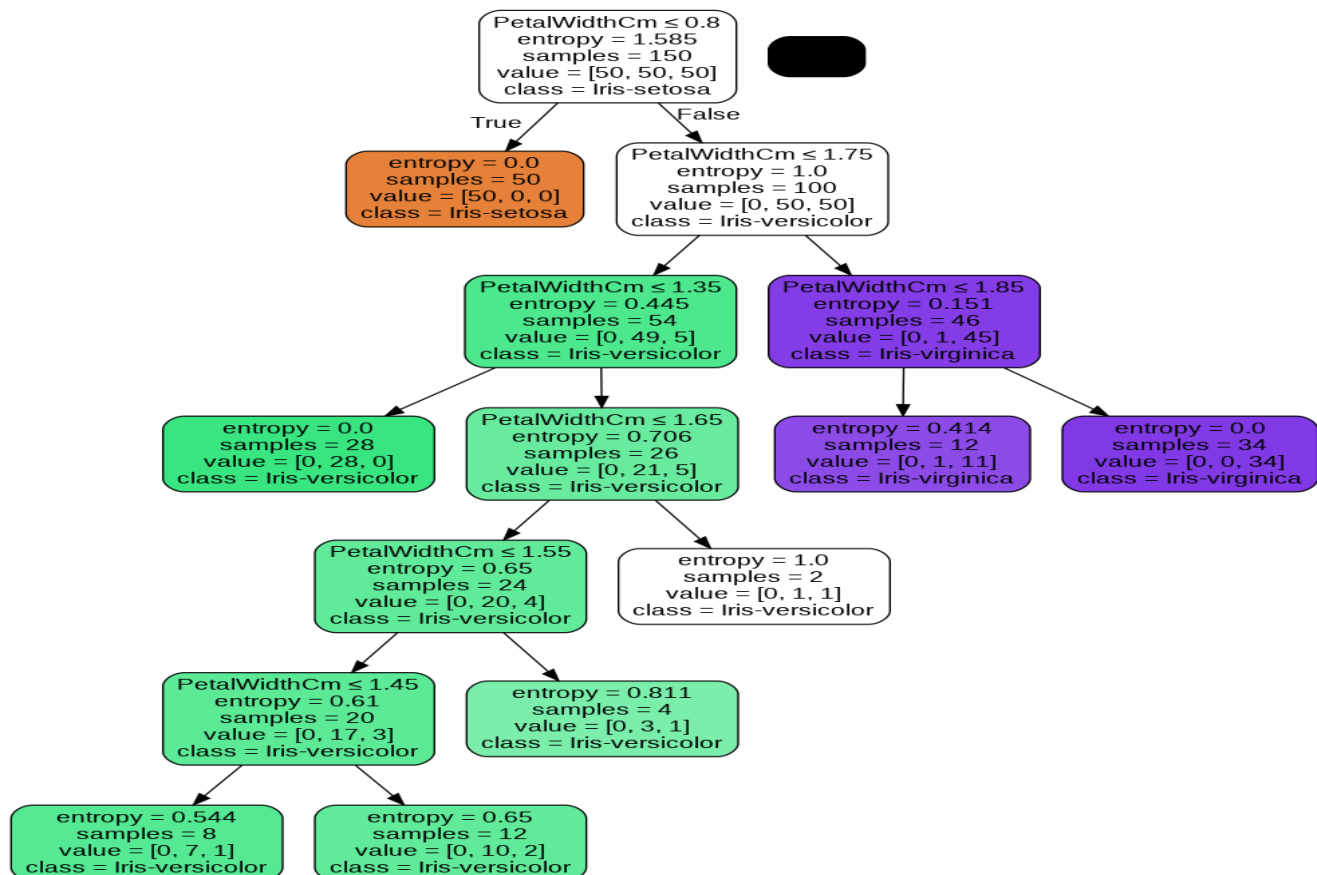
Node has 50 elements of Class 1  
Node has 47 elements of Class 2  
Entropy of Left Node 1.5843574128205722

Right Node  
Node has 3 elements of Class 2  
Entropy of Right Node 0.0

Info gain at split 2.45 is 0.03229223615699528

```
print(info_gains.index(max(info_gains)))
split_values[info_gains.index(max(info_gains))]
clf = tree.DecisionTreeClassifier(criterion='entropy')
clf = clf.fit(X_train,Y_train)
plot_decision_tree(clf, X_train.columns, ["Iris-setosa", "Iris-
versicolor", "Iris-virginica"])
```

5



## 1.7 Results and Discussion

To Implement ID3 (information gain) algorithm for decision tree learning for transforming continuous variables into discrete variables done successfully by calculating Entropy and Information Gain.

