Date: 11/09/2022

Assignment No. 03

Title: Decision Tree Classifier

NAME: Saurabh Nagre

PRN: 2019BTECS00080

**Objective / Aim:** Perform classification task using decision tree, and measure the metrics related to confusion matrix.

**Introduction**:

Decision tree algorithm is a supervised learning algorithm that can be used in both classification and regression analysis. Unlike linear algorithms, decision trees algorithms are capable of dealing with nonlinear relationships between variables in the data.

**Theory / Algorithms:**

**Information Gain:**

To measure the information gain we use the entropy, Which is a quantified measurement of the amount of uncertainty because of any process or any given random variable.

Information Gain = 1 – Entropy

https://149695847.v2.pressablecdn.com/wp-content/uploads/2021/11/image-6.png

Where

H(X) = entropy of variable X

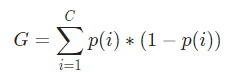
X = random variable or process

Xi = possible outcomes

p(Xi) = probability of possible outcomes.

**Gini Index:**The Gini Index or Gini Impurity is calculated by subtracting the sum of the squared probabilities of each class from one.

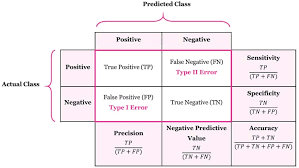
 Gini Index, the formula is given by



Where, C is the total number of classes and p(*i*) is the probability of picking the data point with the class *i.*

**Confusion matrix:**

A confusion matrix is a chart or table that summarizes the performance of a classification model or algorithm for machine learning processes.

****

**The misclassification rate:**

It shows how often your confusion matrix is incorrect in predicting the actual positive and negative outputs. Find this value by adding the false positive and negative values together and dividing this sum by the total number of values in your data set.

**Procedure:**

**Calculate Information Gain**

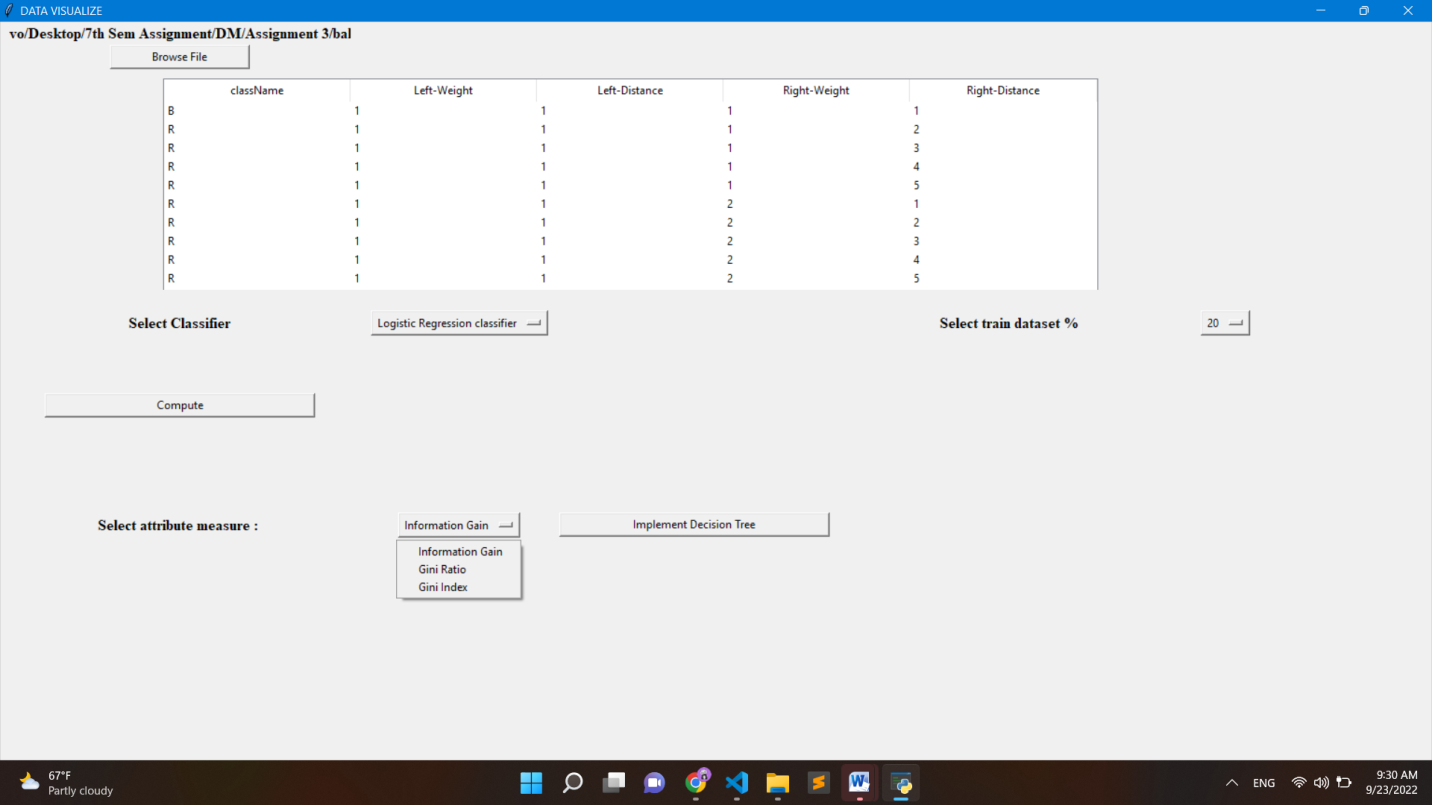
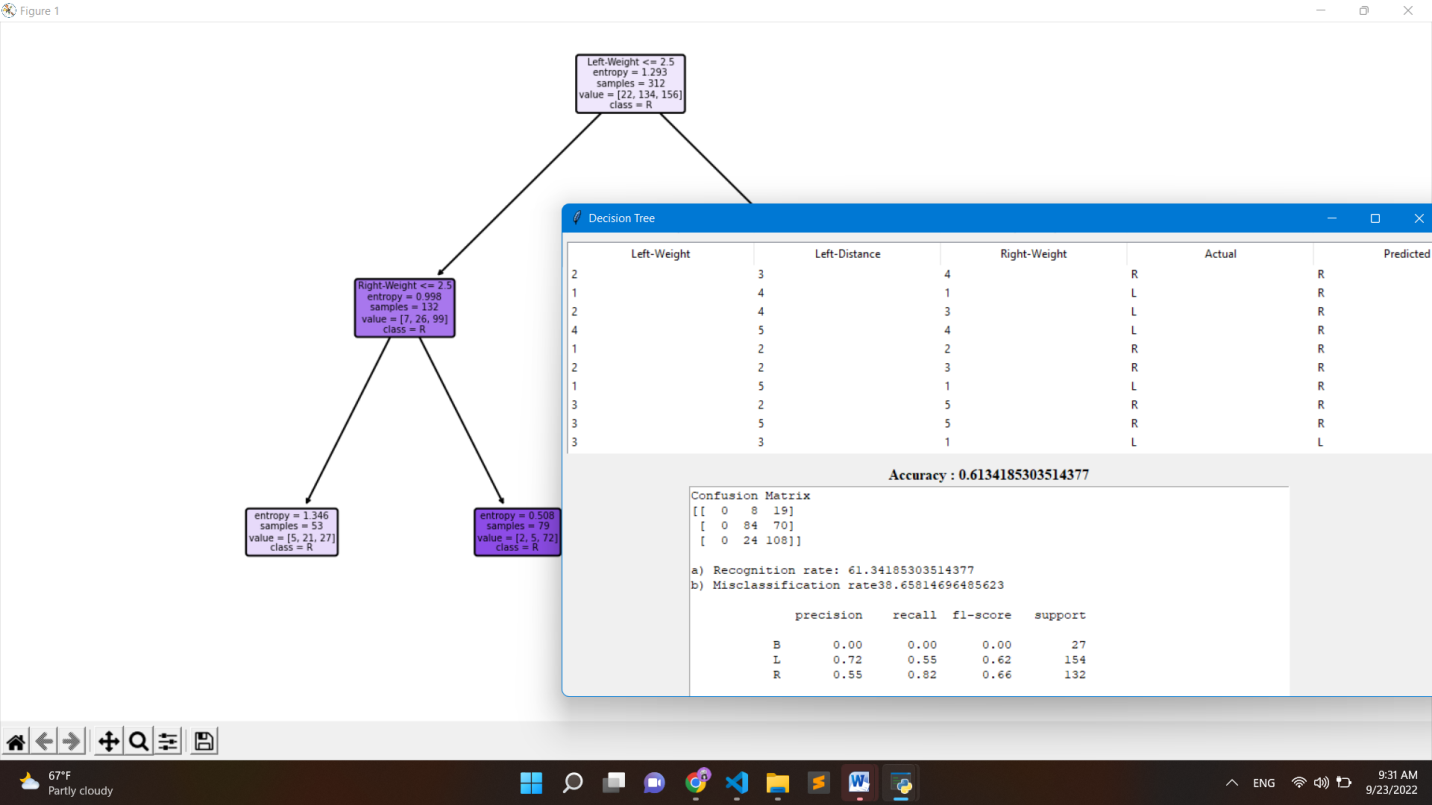
**Design Decision Tree**

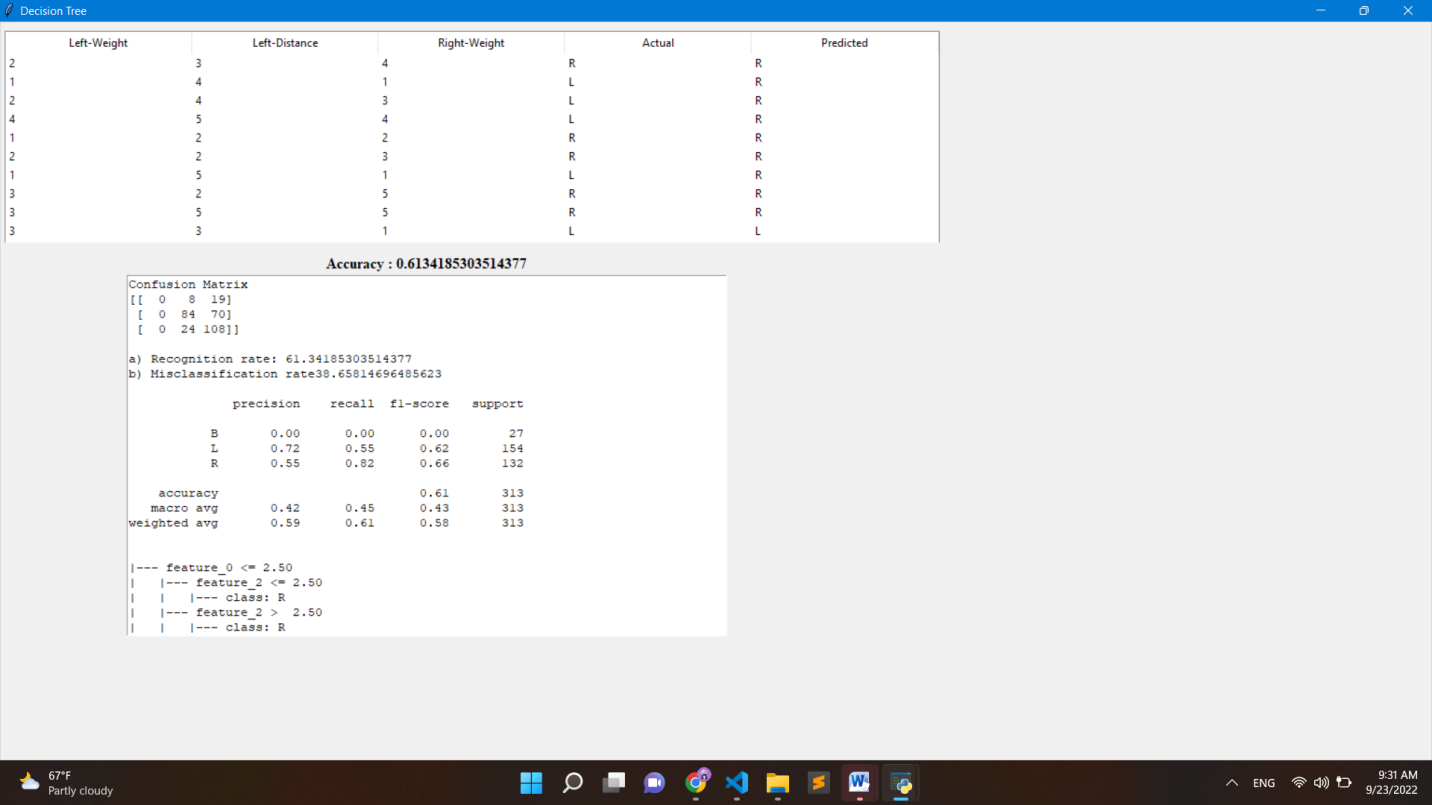
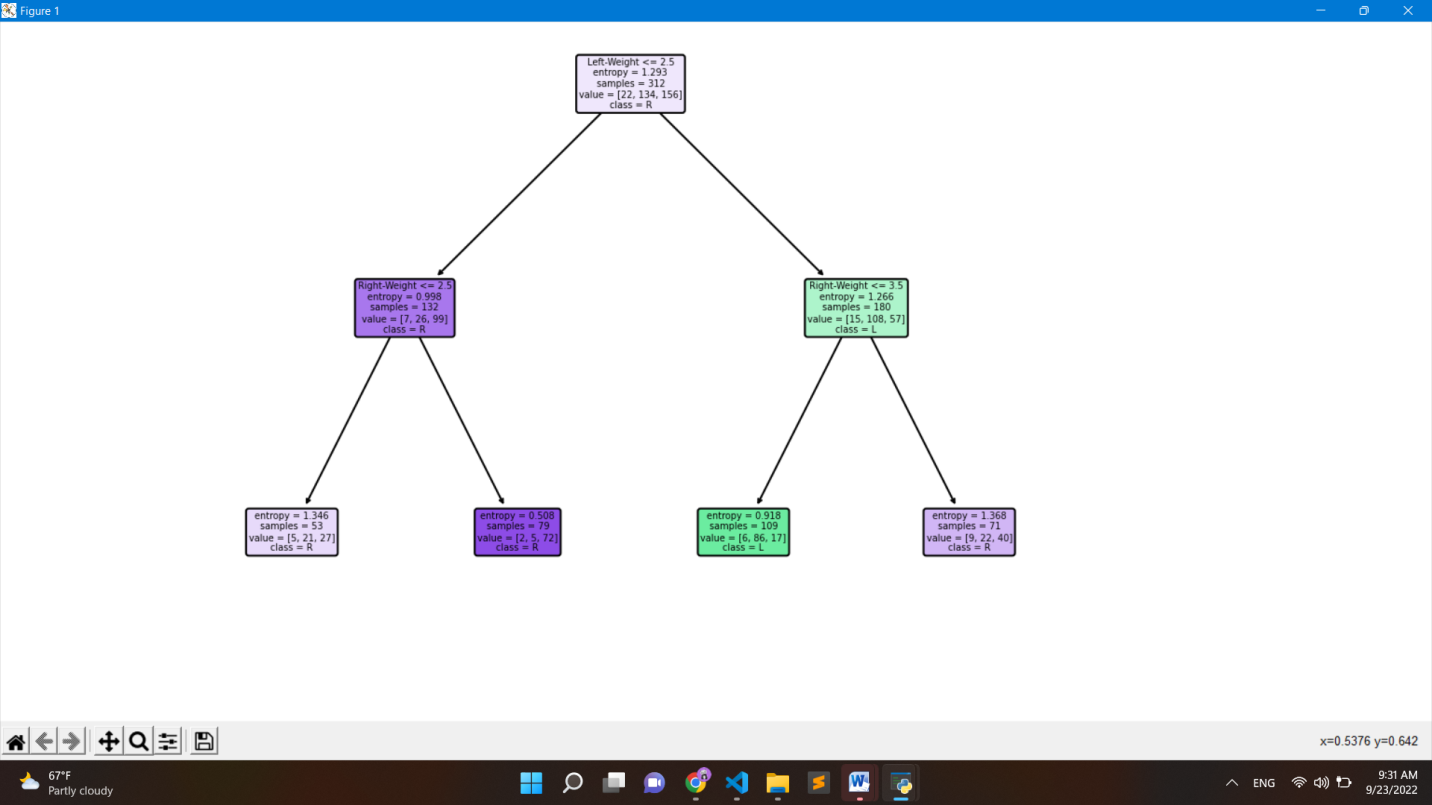
**Calculate Confusion Matrix and**

**Calculate respective formulae’s**

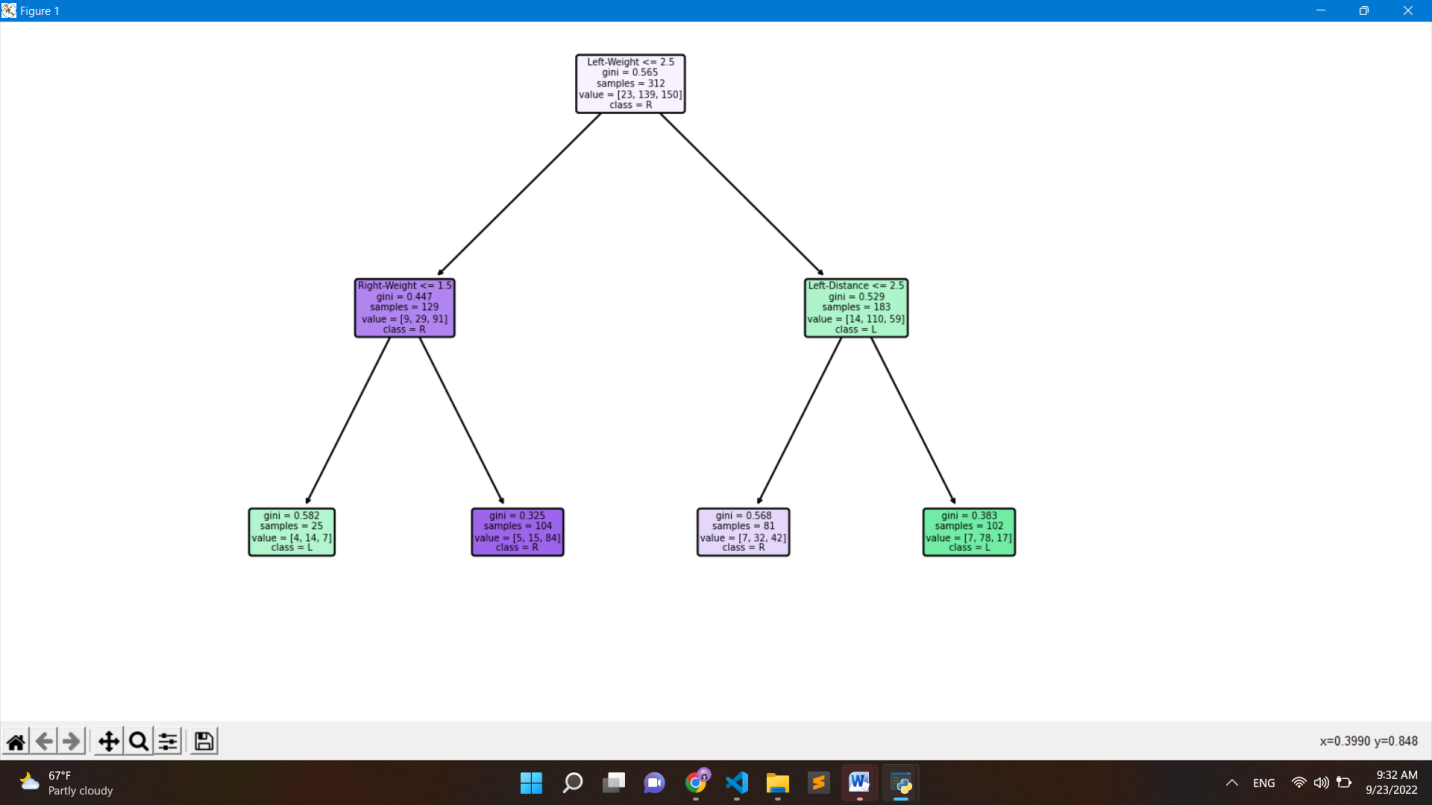
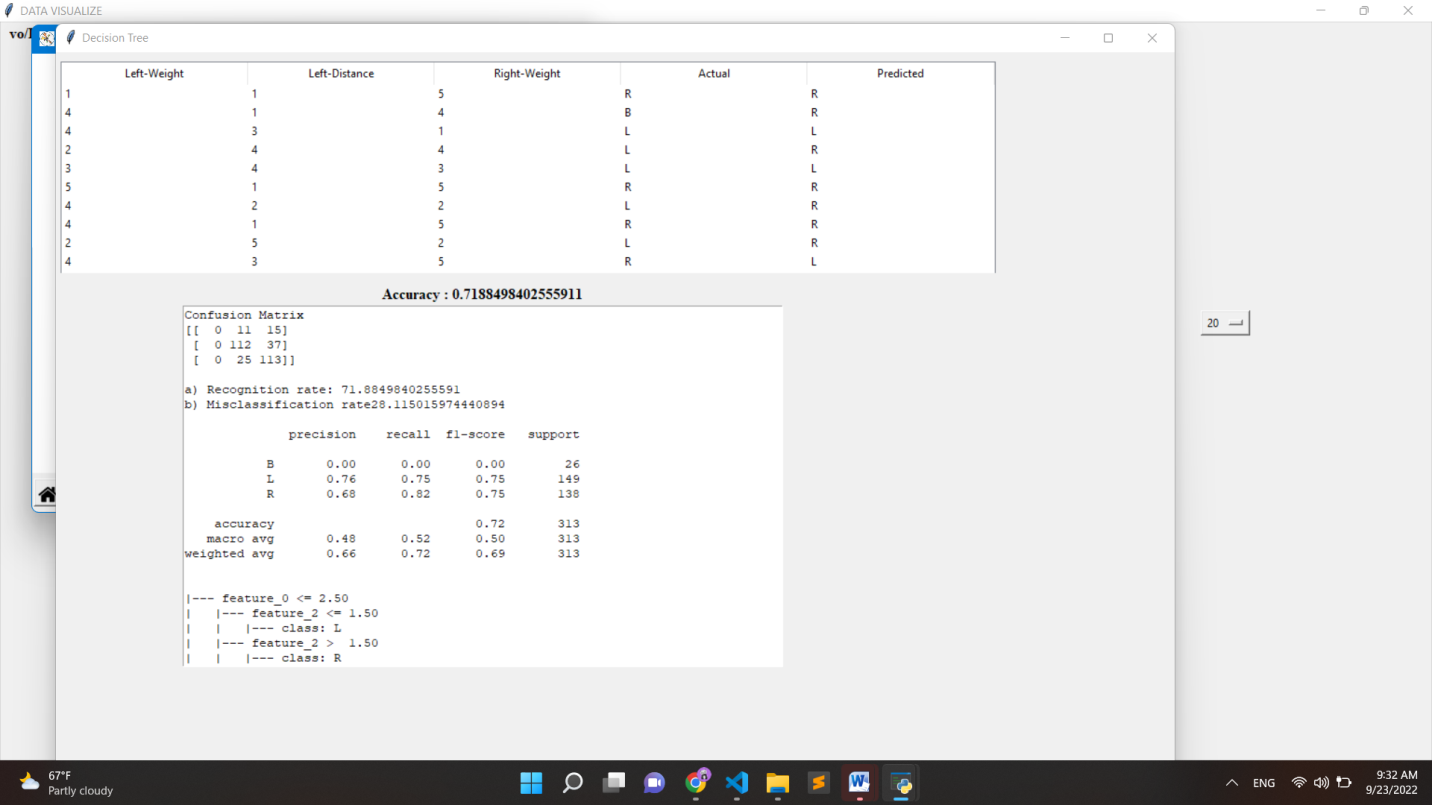
**Actual Experimentation/ simulation/ result/ Observation:**

The GUI:

Information Gain Splitting:



Gini Index Based Splitting:



**Conclusion:**

Implemented Decision tree, and after the predicted result, calculated the confusion matrix. Confusion matrix helped to find out accuracy of the model.

**Reference:**

[**https://www.upgrad.com/blog/gini-index-for-decision-trees/**](https://www.upgrad.com/blog/gini-index-for-decision-trees/)

[**https://www.indeed.com/career-advice/career-development/confusion-matrix#:~:text=The%20misclassification%20rate%20shows%20how,values%20in%20your%20data%20set**](https://www.indeed.com/career-advice/career-development/confusion-matrix%23:~:text=The%20misclassification%20rate%20shows%20how,values%20in%20your%20data%20set).