## **LAB Assignment No. 3**

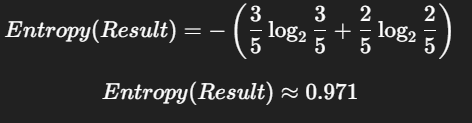
## **Question 1: Entropy and Information Gain (Manual Calculation)**

### **Given Dataset**

|  |  |  |  |
| --- | --- | --- | --- |
| **Student** | **Study Hours** | **Attendance** | **Result** |
| S1 | Low | Poor | Fail |
| S2 | High | Good | Pass |
| S3 | High | Poor | Pass |
| S4 | Low | Good | Fail |
| S5 | High | Good | Pass |

### **1. Entropy of Target Variable (Result)**

Total records = 5  
 Pass = 3  
 Fail = 2



### **2. Information Gain for Study Hours**

#### **Study Hours = High**

* Pass = 3
* Fail = 0

Entropy = 0

#### **Study Hours = Low**

* Pass = 0
* Fail = 2

Entropy = 0

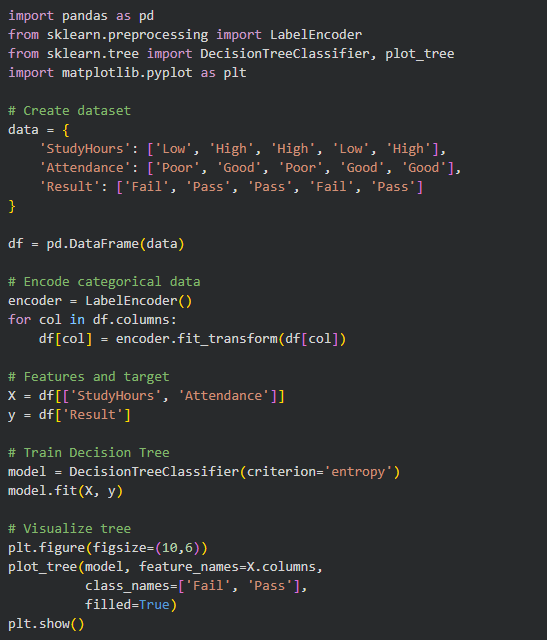
IG(StudyHours)=0.971−0=0.971

### **3. Root Node Selection**

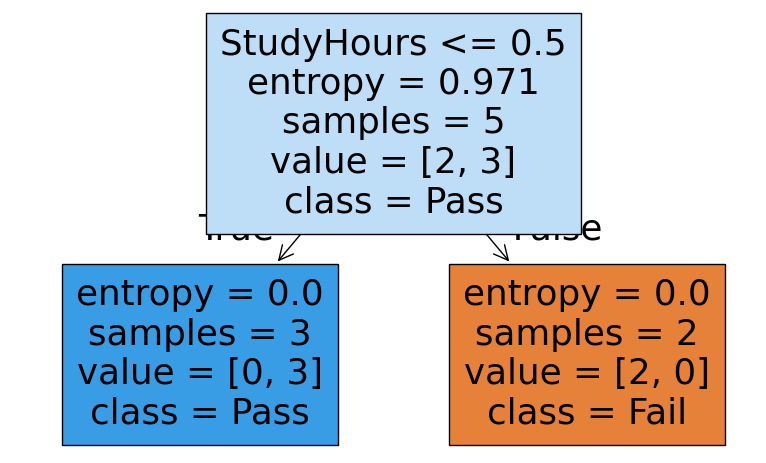
✅ **Study Hours** has the **maximum information gain**, so it should be selected as the **root node**.

## **Question 2: Decision Tree on Small Dataset (Python Implementation)**

### **Python Code**



### **Output**



### **Explanation**

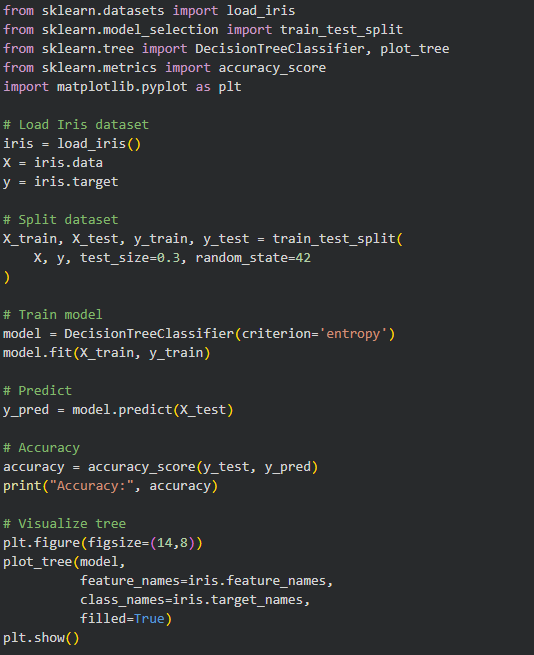
* Data is converted into numeric form using LabelEncoder
* Entropy is used to split nodes
* Tree visualization shows decision rules clearly

## **Question 3: Decision Tree Classifier on Iris Dataset**

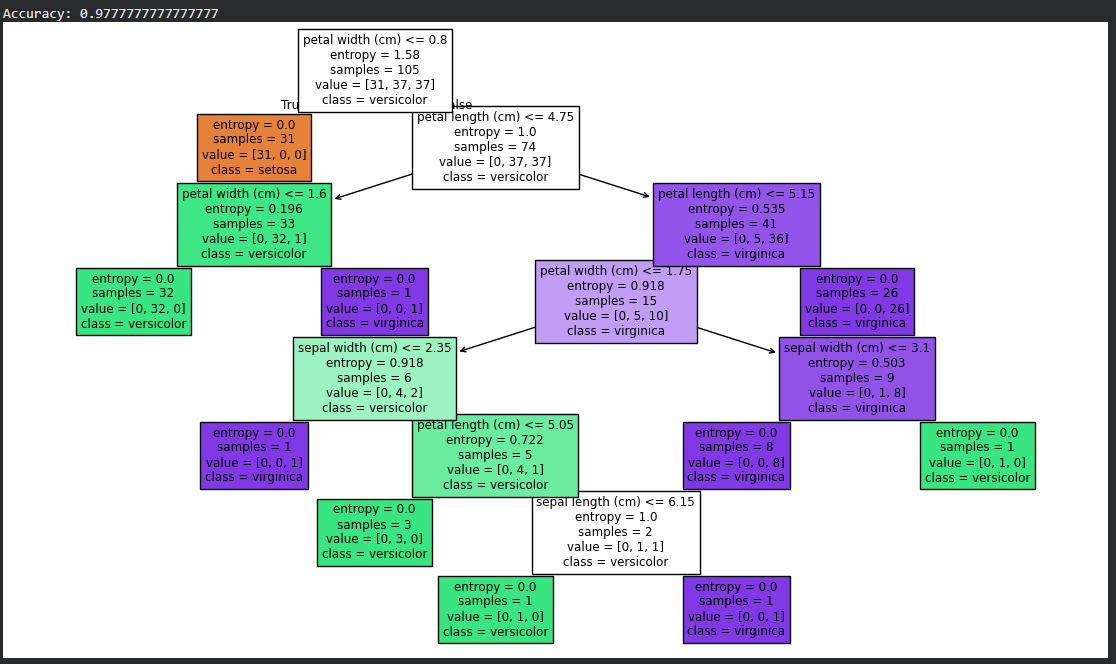
### **Objective**

Apply Decision Tree on a real-world dataset and evaluate accuracy.

**Python Code**



### **Output**

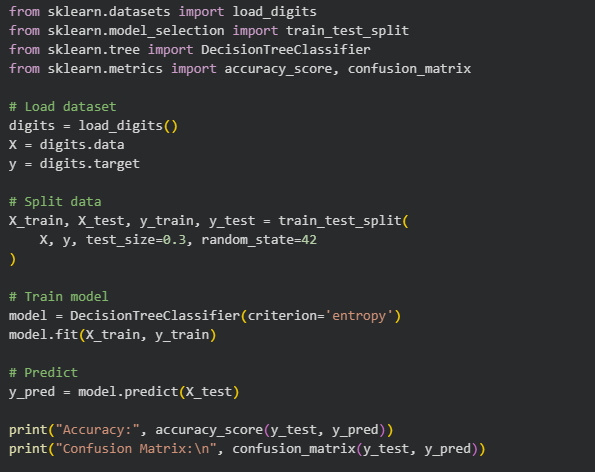


## **Question 4: MNIST Digit Dataset – Baseline Study**

### **Objective**

* Preprocess image data
* Train Decision Tree
* Evaluate performance

### **Python Code**

  
**Output**



### **Discussion & Limitations**

* Decision Trees work on MNIST but are **not ideal** for high-dimensional image data
* They overfit easily
* Deep learning models perform better on images