Exercise 5-1: Decision Tree

# Introduction

This exercise aims to strengthen your skills of building a decision tree through the calculation of Gini indices.

# Question 1

Calculate the decision tree for the following data set. There are four attributes, namely “Hair”, “Height”, “Weight” and “Lotion”. The class attribute is “Result”.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Hair | Height | Weight | Lotion | Result |
| Blonde | Average | Light | No | *Sunburned* |
| Blonde | Tall | Average | Yes | *None* |
| Brown | Short | Average | Yes | *None* |
| Blonde | Short | Average | No | *Sunburned* |
| Blonde | Average | Light | No | *None* |
| Brown | Tall | Heavy | No | *None* |
| Brown | Average | Heavy | No | *None* |
| Blonde | Average | Light | No | *None* |

1. Calculate the Gini index for the attribute “Lotion”.
2. Calculate the Gini index for the attribute “Hair”.
3. Given that the Gini index for the attributes “Height” and “Weight” are 0.3125 and 0.3333 respectively. Which attribute should the root of the decision tree take?
4. Complete the decision tree and draw the tree below.

# Question 2

Consider the following dataset about movie recommendation. Build a decision tree to solve the classification problem. Justify your choice of attribute to split at each node of the decision tree.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Movie ID | Age Group | Genre | Audience Size | Recommendation |
| 1 | Adult | Drama | ok | *No* |
| 2 | ≥12 | Comedy | high | *Yes* |
| 3 | ≥15 | Comedy | low | *No* |
| 4 | ≥15 | Drama | ok | *Yes* |
| 5 | ≥15 | Drama | very-low | *No* |
| 6 | ≥12 | Drama | very-low | *Yes* |