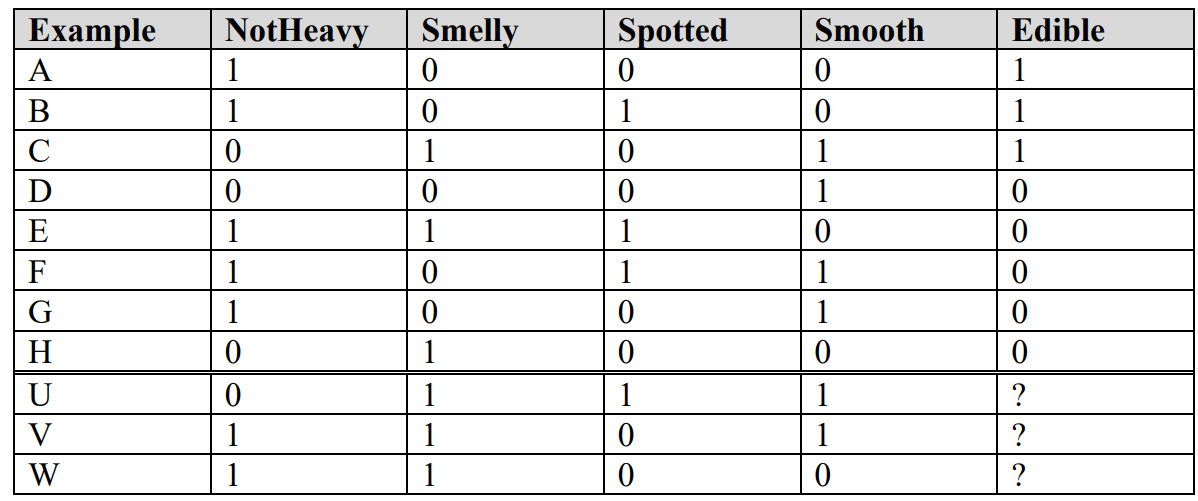
**Introduction to Artificial Intelligence**

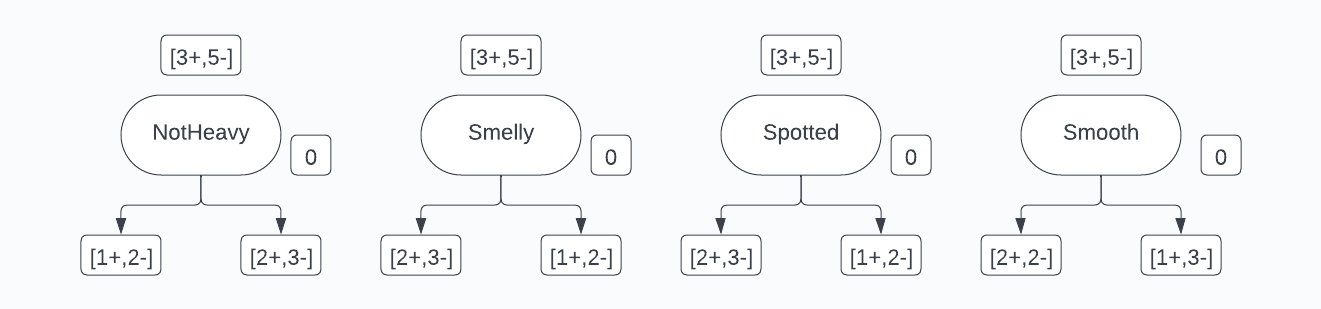
Quiz 7 – Naïve Bayes Classifier & ID3 tree

1. **Problem 1:**

You are stranded on a deserted island. Mushrooms of various types grow widely all over the island, but no other food is anywhere to be found. Some of the mushrooms have been determined as poisonous and others as not (determined by your former companions’ trial and error). You are the only one remaining on the island. You have the following data to consider:



You know whether or not mushrooms A through H are poisonous, but you do not know about U through W.

1. Build a ID3 decision tree to classify mushrooms as poisonous or not.

HEdible

H0/Smooth

IG0/Smooth = HEdible - H0/Smooth = - = 0.048

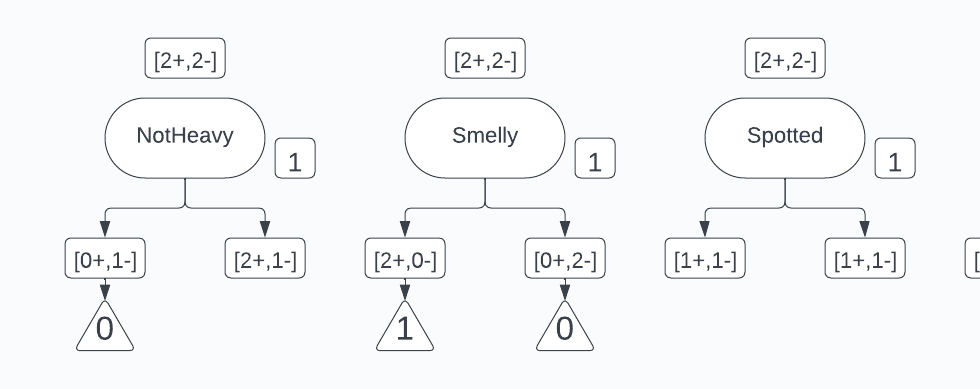
H0/NotHeavy

=> IG0/NotHeavy = HEdible - H0/NotHeavy = - = 0.032 = 0.048

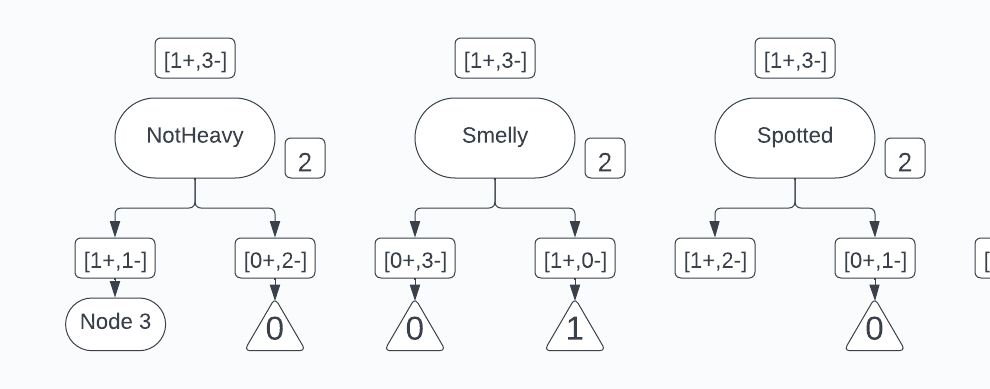
IG0/NotHeavy = IG0/Smelly = IG0/Spotted = 0.032 < IG0/Smooth = 0.048

=> Choose Smooth as root node

If node 1: Smooth = 0

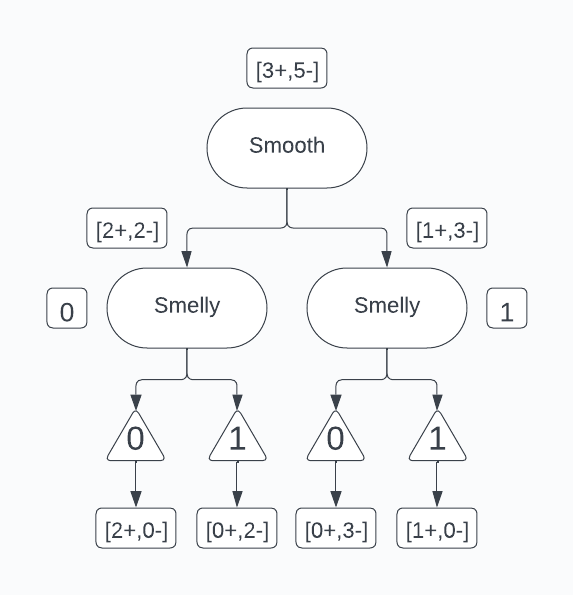


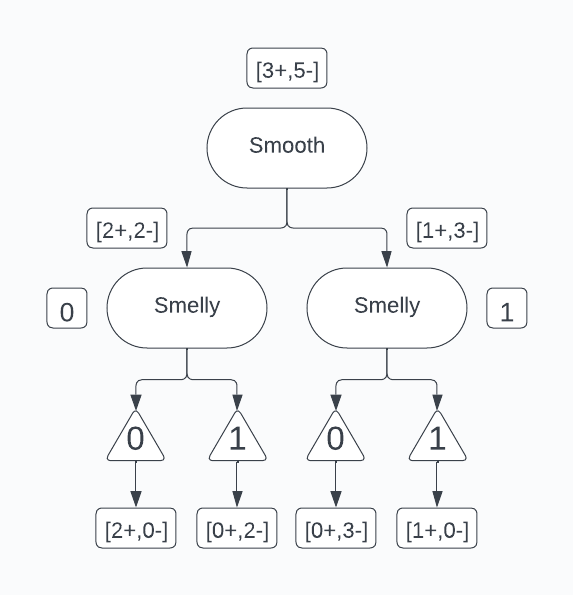
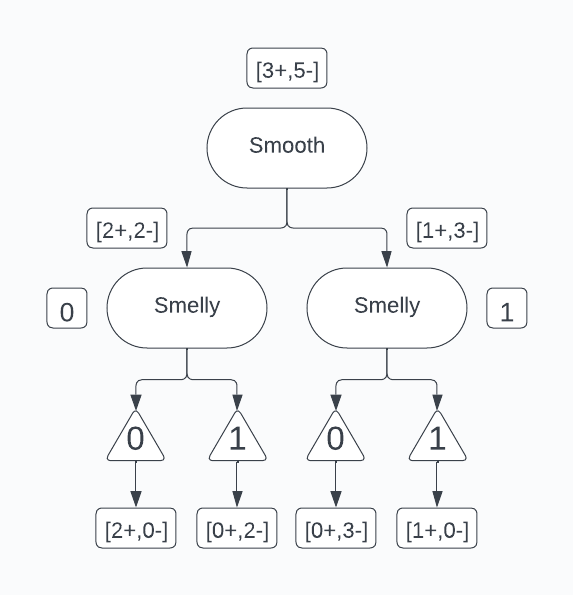
If node 2: Smooth = 1



=> Choose Smelly as second node

The resulting ID3 tree:





1. Classify mushrooms U, V and W using the decision tree as poisonous or not poisonous.

U: Smooth = 0, Smelly =1 => Edible = 1

V: Smooth = 1, Smelly =1 => Edible = 1

W: Smooth = 0, Smelly =1 => Edible = 1

U and V both classify as not poisonous

W classifies as poisonous

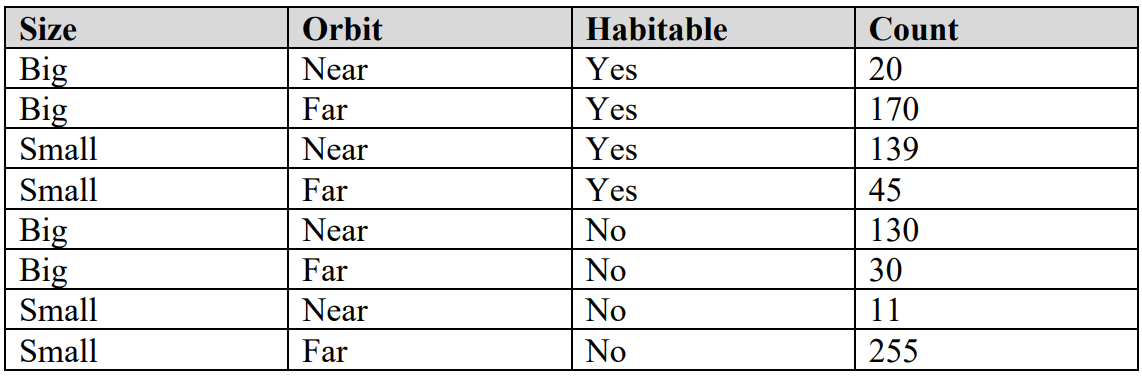
1. If the mushrooms A through H that you know are not poisonous suddenly became scarce, should you consider trying U, V and W? Which one(s) and why? Or if none of them, then why not?

If the mushrooms A through H that are not poisonous suddenly became scarce, I would consider trying W based on the decision tree.

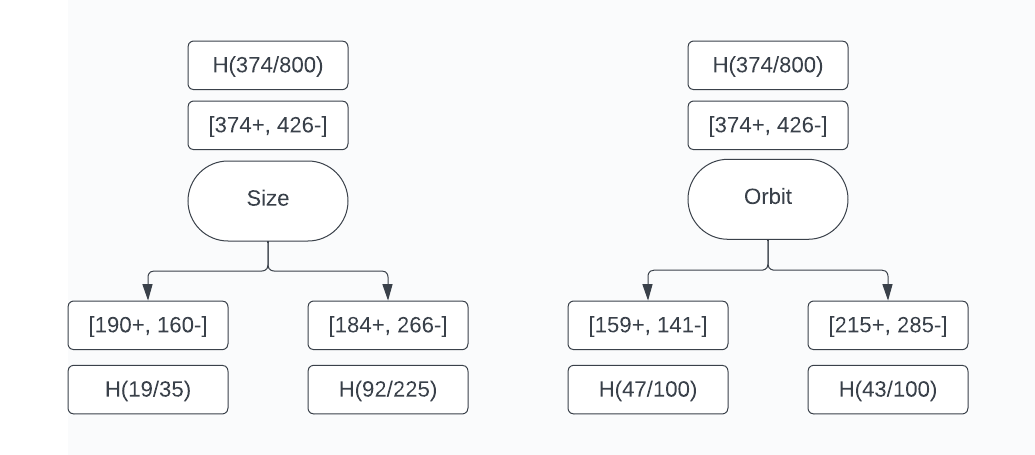
1. **Problem 2:**

As of September 2012, 800 extrasolar planets have been identified in our galaxy. Supersecret surveying spaceships sent to all these planets have established whether they are habitable for humans or not, but sending a spaceship to each planet is expensive. In this problem, you will come up with decision trees to predict if a planet is habitable based only on features observable using telescopes.

1. In below table you are given the data from all 800 planets surveyed so far. The features observed by telescope are Size (“Big” or “Small”), and Orbit (“Near” or “Far”). Each row indicates the values of the features and habitability, and how many times that set of values was observed. So, for example, there were 20 “Big” planets “Near” their star that were habitable



Derive and draw the decision tree learned by ID3 on this data. Make sure to clearly mark at each node what attribute you are splitting on, and which value corresponds to which branch. By each leaf node of the tree, write in the number of habitable and inhabitable planets in the training data that belong to that node



HHabitable

If the mushrooms A through H that are not poisonous suddenly became scarce, I would consider trying W based on the decision tree.

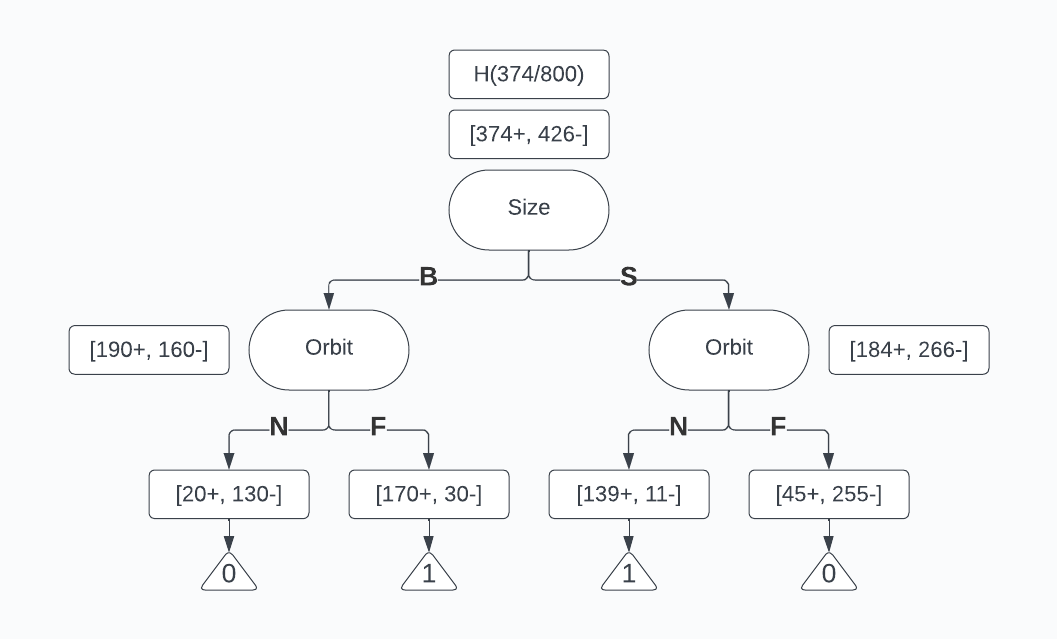
H(Habitual/Size)

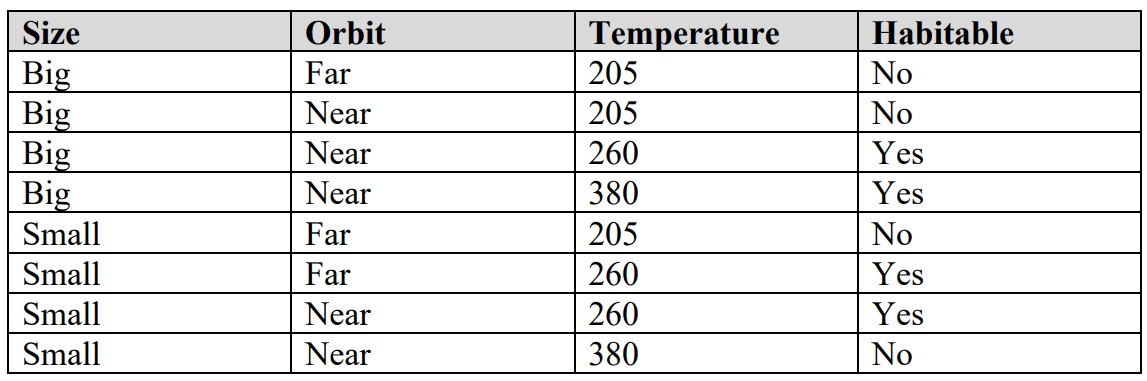
H(Habitual/Orbit)

IG(Habitual/Size)

IG(Habitual/Orbit)

IG(Habitual/ Orbit) = < IG(Habitual/Size) = 0.

Therefore, we choose size as root node, the ID3 decision tree will become:

2. For just 9 of the planets, a third feature, Temperature (in Kelvin degrees), has been measured, as shown in the nearby table. Redo all the steps from part 1 on this data using all three features. For the Temperature feature, in each iteration you must maximize over all possible binary thresholding splits (such as T ≤ 250 vs. T > 250, for example).

The 9th row is small, near, 380 and no

Binary threshold splits for the continuous attribute temperature

205oC: [0+, 3-] < 232.5oC < 260oC: [3+, 0-] < 320oC <380oC: [1+, 2-]

HHabitable

H(Habitable/Size)

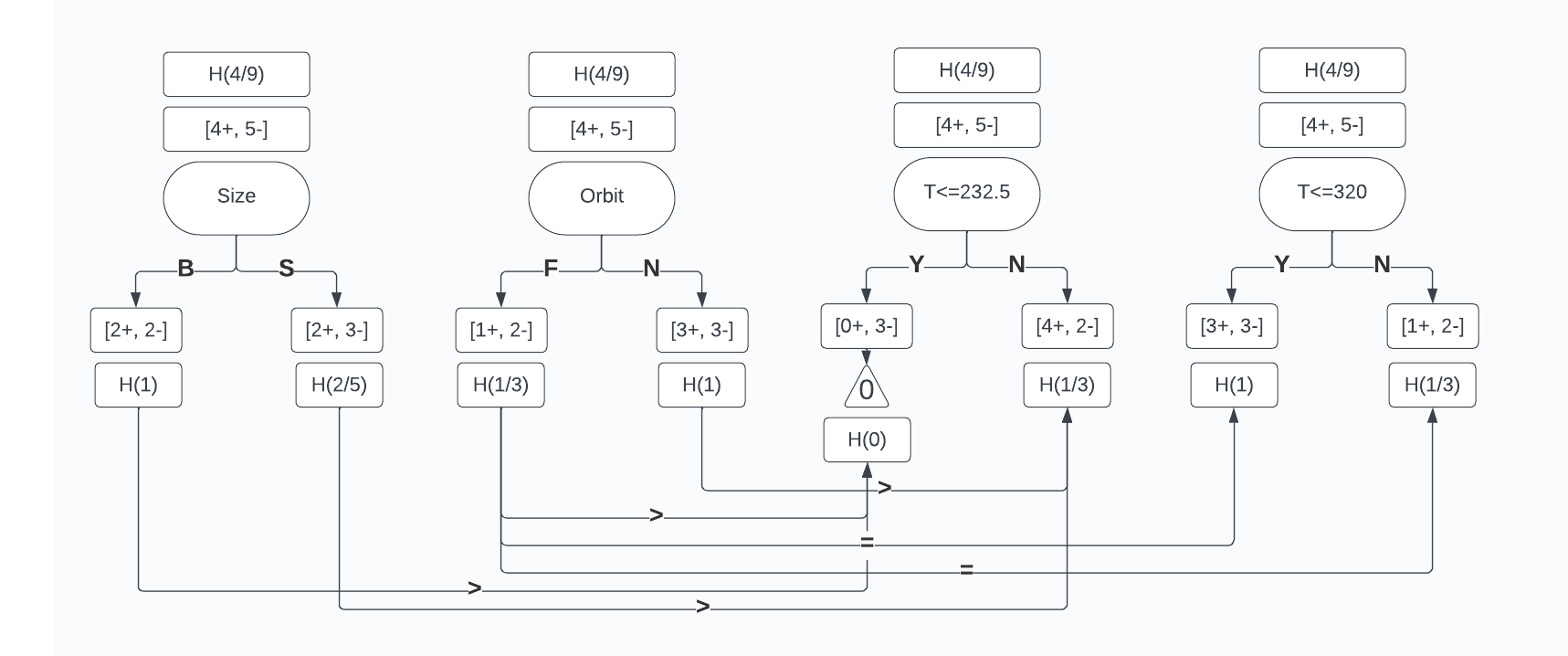
H(Habitable/Temp<=232.5)

H(Habitable/Orbit)

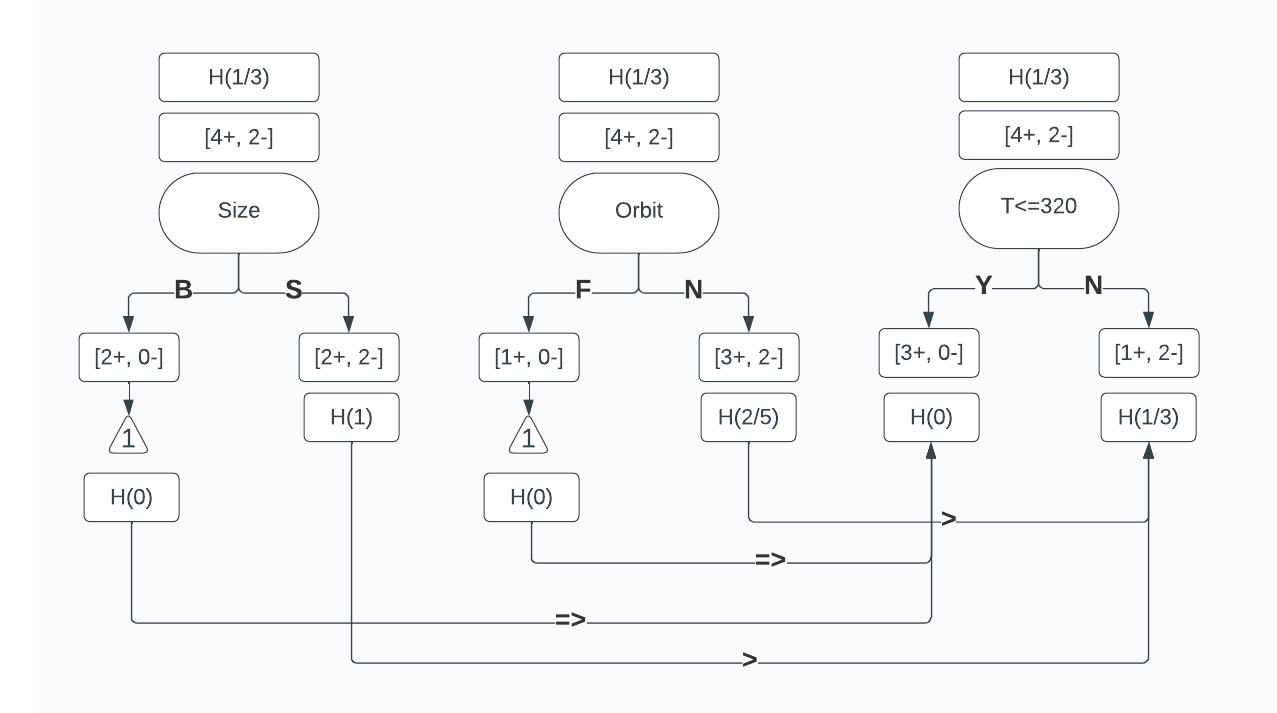
H(Habitable/Temp<=320)

IG(Habitable/Size)

IG(Habitable/Orbit) = IG(Habitable/ Temp<=320)

IG(Habitable/Temp<=232.5)  > IG(Habitable/Size) > IG(Habitable/ Orbit) = IG(Habitable/ Temp<=320)

Thus, we choose (T <= 232.5) as root node



HTemp<=232.5

H(Temp<=232.5/Size)

H(Temp<=232.5/Temp<=320)

H(Temp<=232.5/Orbit)

IG(Temp<=232.5//Temp<=320)

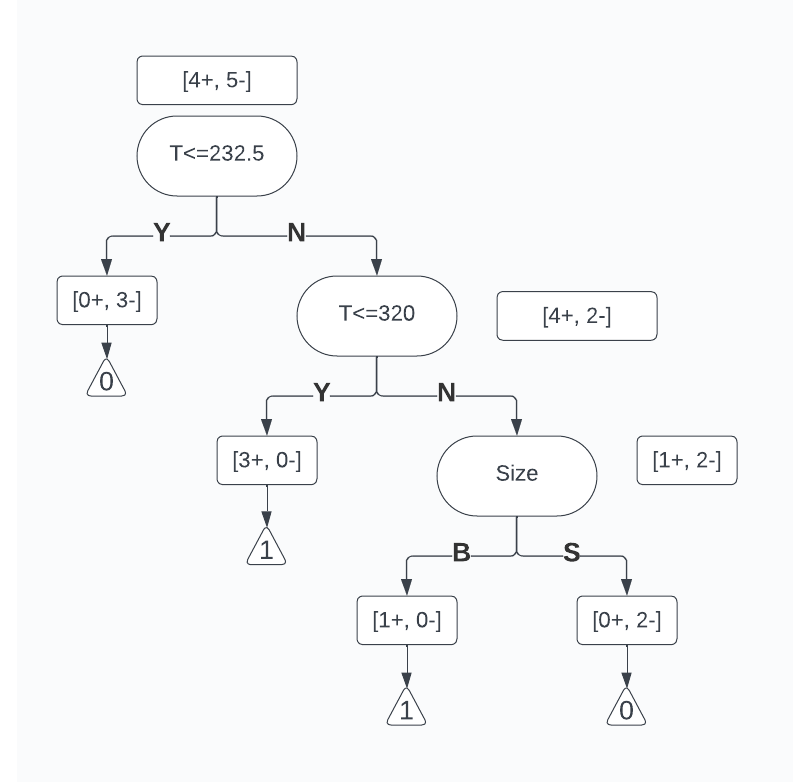
IG(Temp<=232.5/Orbit) =

IG(Temp<=232.5/ Size) =

IG(Temp<=232.5/Temp<=320)  > IG(Temp<=232.5/Size) > IG(Temp<=232.5/Orbit)

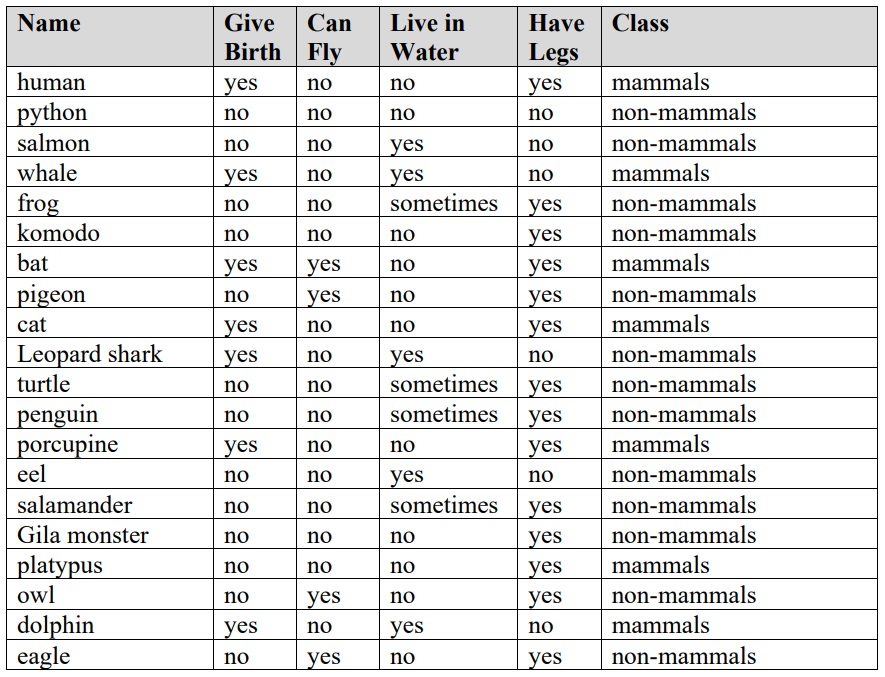
Therefore the level 2 will become T <=320oC

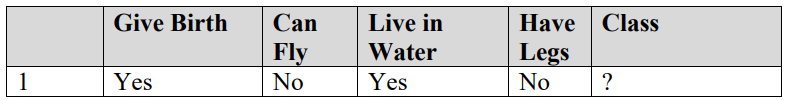
So the final decision tree is:



According to my decision tree, the planet with the features (Big, Near, 280) is habitable.

1. **Problem 3:**

Given dataset about animal

Using Naïve Bayes Classifier to find whether this animal is mammal or not (Using Laplacian smoothing if need)

Pmammal ; Pnon-mammal

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Give birth | Can fly | Live in water | Have legs | Total |
|  | Yes | No | Yes | No |  |
| Mammal | 6 | 6 | 2 | 2 | 7 |
| Non-mammal | 1 | 10 | 3 | 4 | 13 |

P(Mammal|X) = P(X|Mammal) x P(Mammal) = P(GiveBirth|Mammal) x P(CanFly|Mammal) x P(LiveInWater|Mammal) x P(HaveLegs|Mammal) x P(Mammal)

P(Non-mammal|X) = P(X|Non-mammal) x P(Non-mammal) = P(GiveBirth|Non-mammal) x P(CanFly|Non-mammal) x P(LiveInWater|Non-mammal) x P(HaveLegs|Non-mammal) x P(Non-mammal)

Because P(Mammal|X) > P(Non-mammal|X), we conclude that the animal is mammal